DOCUMENT RESUME

ED 160 [illegible]2 SP 012 821

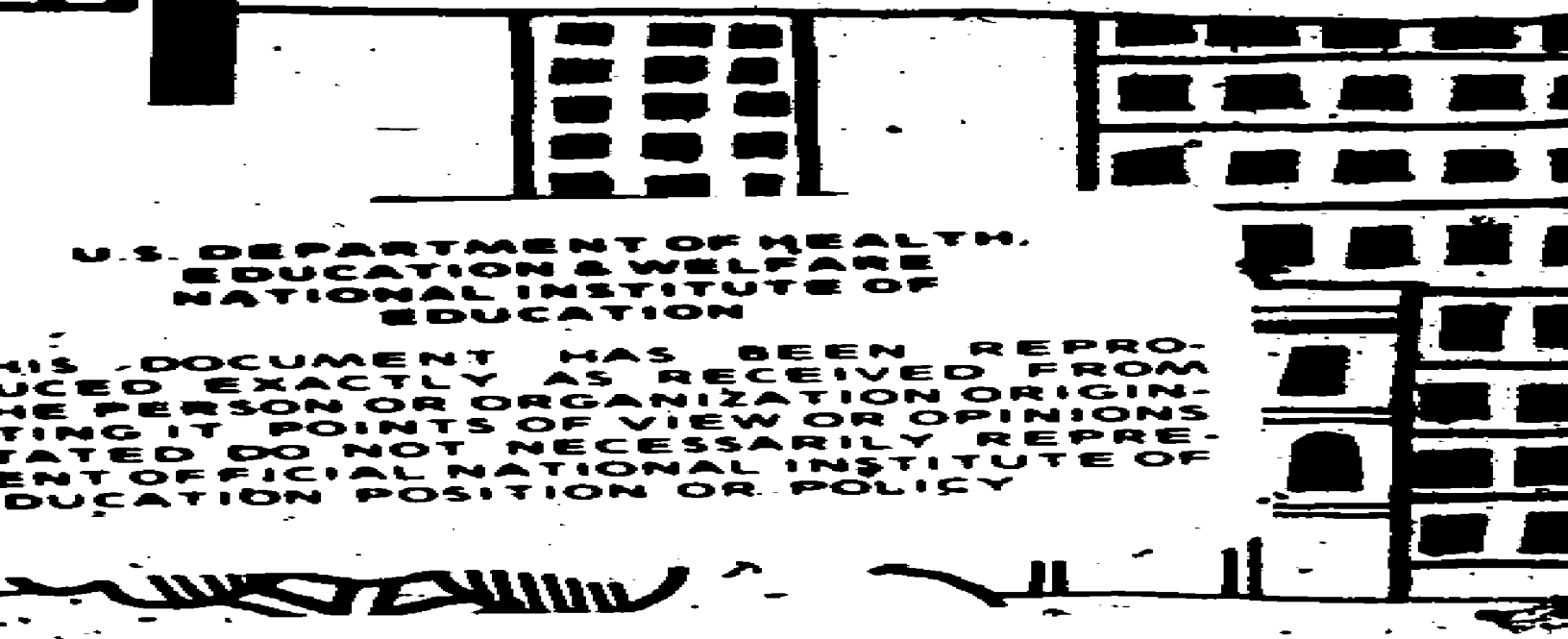
AUTHOR Whitehead, Graham
TITLE Enquiry Learning in Social Studies. ACER Research Series No. 101.
INSTITUTION Australian Council for Educational Research, Hawthorn.
PUB DATE Apr 78
NOTE 180p.
AVAILABLE FROM Australian Council for Educational Research Limited, P. O. Box 210, Hawthorn, Victoria, Australia 3122

EDRS PRICE MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS Child Development; Cognitive Development; *Critical Thinking; Educational Research; Elementary School Students; *Experimental Programs; Grade 6; *Inquiry Training; *Learning Processes; Methods Teachers; Primary Education; Problem Solving; Program Descriptions; *Social Studies; Statistical Studies; *Teaching Techniques

ABSTRACT

Teaching enquiry methods, that is, skills concerned with constructing knowledge from available information and searching for new information, is the subject of this study, which provides a background to the subject as well as a review of research. The study sought to answer the question of whether problem-solving performances of primary school children could be improved through training. Two training programs were developed and implemented with 216 grade-six children, drawn from six primary schools located in middle-class areas of Melbourne, Australia. Attention is given to the rationale behind the training, the nature of the training, the evaluation techniques, and the experimental design employed. Results showed that those children in the investigation who had the benefit of a teacher who taught the skills and explained what was expected were clearly superior in performance to students whose teacher encouraged them to talk but who primarily acted as discussion controller rather than as an instructor. Tables illustrating the collected data, and references, are included. (DS)

INQUIRY LEARNING SOCIAL STUDIES



graham
whitehead

Enquiry Learning in Social Studies

ACER Research Series No. 101

Published by the Australian Council for Educational Research, Limited,
Frederick Street, Hawthorn, Victoria 3122
April 1978

Typeset by
Meredith Trade Lino Pty Ltd,
Burnley, Victoria 3121

Printed and bound by
Brown Prior Anderson Pty Ltd,
Burwood, Victoria 3125

National Library of Australia Cataloguing-in-Publication data.

Whitehead, Graham J.
Enquiry learning in social studies.

(Australian Council for Educational Research.
Research series; no. 101).
Bibliography.
ISBN 0 85563 164 3.

1. Cognition (Child psychology). 2. Social
sciences — Study and Teaching. I. Title. (Series).

370.152

For Pam, Mark and Anne

Table Of Contents

List of Tables

List of Figures

Foreword

Some of the main issues in current educational debates are concerned with the core curriculum, the learning of skills, and the balance between content and process learning. These are closely related issues, for many intellectual skills, most obviously those of communication and literacy, are vital to everyday activity and are readily seen as part of the core curriculum. It is well recognized too, for example, in programs concerned with language across the curriculum, that such skills are not separate curriculum components but pervasive aspects of all school activities.

In the same way, many would press for the skills associated with thinking, enquiry, and problem-solving to be accorded a more prominent place in the core curriculum, again not as separate units of study but as integral to many or all school activities. However, just as many communication skills need to be taught, so, too, may thinking and enquiry skills. This book is concerned with such teaching.

There have been many attempts at teaching enquiry methods, that is, skills concerned with constructing knowledge from available information and searching for new information. There have been similar attempts with teaching the closely related skills of problem-solving, those concerned with putting available information to use. Indeed, problem-solving is most frequently seen as a form of enquiry. However, the career of such skills as feasible curriculum components has been a chequered one, as Dr Whitehead's review shows. Apart from the problems that generally beset educational innovation, attempts at teaching enquiry skills face at least three important difficulties.

First, there is often uncertainty in a particular area as to the exact nature of enquiry skills or problem-solving. The best way to demonstrate this diffi-

culty is to invite the reader to compose a number of test tasks which would indicate whether a grade six child is demonstrating some agreed level of enquiry skill. It is clear that such skills are much less easy to assess, and even to specify, than the skills of, say, reading and calculation. It is indeed much easier to recognize the solution to a problem or the statement of a principle gained from enquiry than it is to recognize the quality or appropriateness of the steps taken in the process of enquiry itself. For this reason, teaching most frequently focuses on the product rather than the process.

Secondly, it is uncertain whether enquiry skills that are appropriate for one area of knowledge are suitable for others, or whether such skills are general from one aspect of a particular subject to another. The enquiry skills of science for example, are in many ways different from those of social studies. One of the tasks of the primary school would seem to be to identify and teach the common and basic elements of the enquiry process across different content boundaries.

Thirdly, there is the question of how the process of enquiry can best be taught, or indeed whether it can or should be taught at all. The art of enquiry or of solving problems might well be some natural process that matures as one gains more experience of the world, developing in each person according to his native ability. Alternatively, although we may learn to learn and discover how to discover essentially through our own devices, we may develop these skills all the better through systematic opportunities to practise them. The amount of practice may well be the critical factor. On the other hand, just as other skills can be analysed, and efficient teaching programs based on component skills devised, so it may be with the process of enquiry. If this last alternative is the case, there is a clear need to develop these efficient teaching procedures.

It is these three major questions which form the main themes of this book. Not only are they analysed and the literature reviewed from a variety of vantage points, but the experimental study described offers an important contribution to their further analysis in terms of social studies teaching.

There is no further need to comment on the substantive issues raised, but one methodological issue deserves attention. The researcher concerned with precise educational questions has an unenviable task, for so frequently the constraints of school life, rightly, make strict experimental work either impracticable or undesirable, and, in any case, uncontrolled variables usually make interpretation doubtful. Even the weaker demands of correlational studies are often not adequately met, and frequently the only justifiable approach is the detailed case study. However, there are cases in which the

nature of the questions deserves a stricter experimental approach. It is then up to the researcher to exercise both care and ingenuity in carrying through a design which does not do violence to the richness of the constructs or the particularities of situations or to the human factors involved. Here is one of those cases in which such a design seems to have been possible.

December 1977

G. T. EVANS.
Professor of teacher education,
Department of Education,
University of Queensland.

Acknowledgments

In initiating, developing and implementing this research I have had the assistance and encouragement of many people. To them all I express my thanks.

Professor Glen Evans, Professor Millicent Poole and Dr John Keeves provided advice at various stages of the project. Their constructive suggestions concerning design, statistical analysis and the preparation of the report have been greatly appreciated. Mrs Clair Burleigh, Mrs Jan Murdoch and Mr David Crickmore were the teachers who taught the program in schools. Mrs Dawn Allen helped in administering the tests and assessing student responses.

In addition to the above people there was a relatively large group of teachers who helped in various ways. Mr Bill Henderson, Miss Pat Maguire, Mr Stan Oakley, Mr Ken Smith and Mr John Waters allowed me to work in their classrooms when developing the test instruments and the training program. The children in the study were drawn from the classrooms of Mr Murray Cowell, Mr Bob Eames, Mr Frank Fielding, Mr Barry Gasperino, Mr Brett Lewis, Miss Diane McIlwraith, Mr Ken McLeod, Mrs Nora Mackieson, Mr Bart Shattock, Mrs Kay Waghorn, Mr Ken Walker and Mr Keith Williams.

To the National Committee on Social Science Teaching and the Education Department of Victoria I am grateful for financial assistance.

Finally, I express my thanks to the children who participated in the investigation.

1. Introduction

> *. . . the child must learn to cope with a rapidly changing environment. To do this he must be able to view his role and those of others in an objective way via the test of evidence. It is believed that this ultimately will be achieved if the child develops and refines his knowledge about the social world, and learns the process of inquiry through using it, so enhancing his ability to think in a rational manner. These three factors are closely interwoven.*
> (*Society in View*, Victoria, Education Department, 1974:6)

The improvement of thinking as an objective of Social Studies teaching receives strong emphasis today. Such an objective is not new and can readily be traced to Dewey at the beginning of the century and to other writers before him. Dewey (1916:170) wrote, 'The sole direct path to enduring improvement in the methods of instruction and learning consists in centering upon the conditions which exact, promote and test thinking.' This theme underlies many of the new Social Studies programs developed recently both in Australia and overseas (Queensland, Education Department, 1972; Trezise, 1974; Victoria, Education Department, 1974; Western Australia, Education Department, 1974; Crickmore & Leighton, 1975). These programs aimed to help the child think in a systematic and rational way about the world in which he lives. The child is encouraged to solve problems, to engage in the process of enquiry, and to seek reasoned decisions. The learner is helped to ask questions, gather information, seek alternative solutions to problems, evaluate information and reach decisions: all facets of enquiry.

Why this attention to enquiry, to problem-solving behaviour, to rational thinking in Social Studies curricula? Such behaviour is seen as part of the

procedure for coping more effectively with social change. In a society experiencing rapid change, Social Studies educators believe a child should be put in the position of being able to evaluate the theories and conclusions of others without being a slave to them. Each individual should be able to determine his own behaviour and bear responsibility for his own actions. While accepting that non-rational elements are part of our common humanity and that there are other relevant values in education, curriculum developers espouse the view that rational behaviour and procedures enlarge knowledge and provide a way by which individuals may more readily control their own destiny (Fish & Goldmark, 1966; Taba, Durkin, Fraenkel & McNaughton, 1971; Banks, 1973). To this end, Social Studies programs advocate the provision of experiences in the process of enquiry.

Underlying the advocacy of enquiry is the assumption that school experiences can make a contribution to the development of abilities of enquiry. According to *Society in View*, 'a child's intellectual abilities are not fixed but can be enhanced through the use of appropriate learning experiences.' (Victoria, Education Department, 1974:11). Earlier, Taba (1966), in taking a similar position, noted 'the very fact that thinking has figured as an educational objective and the frequency with which its central importance in the educational enterprise has been invoked presumes that the capacity to think can be systematically fostered.' (Taba, 1966:31). From such an assumption, various programs have proceeded to detail teaching strategies and develop materials which the authors believed would achieve the stated objectives. Little research has been done in the area of Social Studies to investigate the assumption.

The present study sought answers to the following general questions. Can students be trained to apply the enquiry process? Will the training transfer to Social Studies-type tasks which differ from those experienced in class? Should improvement be noted, how durable will be the success? To find the answers to these and some subsidiary questions, two training programs were developed and implemented with 216 grade six children, drawn from six primary schools located in middle-class areas of Melbourne.

In the following chapters issues related to the background of this research are examined and the results of the research discussed. Chapter 2 explores the nature of the enquiry process and the ability of primary school children to apply it. What is enquiry? How does it relate to critical thinking? Is there a sequence to enquiry? What is the mastery level of primary school children in handling the processes? The reasons why children should be limited in their ability to use the enquiry process are considered in Chapter 3. The views of

Piaget regarding the factors influencing growth are examined, although ultimately it is to the ideas of Pascual-Leone that most attention is given. Implications underlying the Pascual-Leone model for training in problem-solving are noted. Previous research into problem-solving training in the general area of Social Studies is reviewed in Chapter 4. Attention is given to the rationale behind the training, the nature of the training, the evaluation techniques and the experimental design employed. In Chapter 5 the activities used in the present investigation are described. The nature of the problems used in the training sessions, the materials upon which the problems are based and the teaching strategies employed are explored. The design, sampling and experimental method are presented in Chapter 6. Particular attention is paid to describing the tests used, the methods of scoring the responses, and the sample of subjects to whom the tests are administered. The proposed analysis of data is presented together with a statement of the issues which form the foci of the investigation. In Chapter 7 the results of the analyses are reported. In the main, the results of the treatments are discussed in Chapter 8, although, where significant findings related to teacher, school or class effects are noted, these are discussed in Chapter 7. Consideration is given in Chapter 8 to the contribution of the special training programs to performance on the various measures. Observed differences between the treatment groups are related back to the Pascual-Leone model from which the training procedures have been derived. Questions related to durability and transfer are also considered at that point. In Chapter 9 the present investigation is reviewed, and the findings, their educational significance and implications for further research noted.

2. The Enquiry Process and the Child

The Enquiry Process

A major resurgence of interest in the process of enquiry occurred in the context of the natural science curriculum. Bruner (1960) wrote *Process of Education* as a result of the deliberations by a group of people interested in the teaching of science. In this book Bruner took as one of his themes the processes used by students in solving problems. He saw the development of such processes as being an important educational end. He wrote of intuitive thinking as being part of this procedure and stressed the need for school curricula to foster its development. Unfortunately he was somewhat vague in defining its properties, although he did contrast it with analytic thinking, where the thinker tackles a problem in a step by step fashion in accordance with a set plan of attack. In intuitive thinking the individual may leap about with his ideas, employ short cuts and be unaware of how he obtained his answers, yet he will be constantly drawing upon his understanding of a particular body of knowledge. It was this mental leaping that Bruner saw as being crucial in the generation of answers to given problems. Once answers were gained, then Bruner advocated a checking procedure utilizing what he entitled 'analytic methods'. He did not define these methods clearly.

In a much earlier statement on enquiry, Dewey (1933) called for school programs to develop the reflective thinking ability of the child. In the act of reflective thinking he identified five phases. The first centred on the need for direct personal experience of situations drawn from ordinary life, outside the school environment: something new, uncertain or perplexing which caused reflection. Out of this phase grew the question that had to be answered. The second phase was the location and definition of the problem. The third was where the thinker decided what to do with the problem. Here he must have ideas. These may arise from accumulated knowledge but also from infer-

ence, projection, invention or ingenuity. Old ideas emerged in new contexts. From these ideas the child generated possible solutions. These in turn may be modified into definite suppositions or hypotheses. The fourth phase, labelled 'reasoning in the narrow sense' by Dewey, concerned the mental manipulation of the data, relating ideas to the particular problem. There may be a sequence of activity where one idea is linked to another, stretched and elaborated, to form a new idea. The fifth phase was where the proposed solution to the identified problem was verified through application to a new situation. It was in this phase that the doubt originally raised in phase one was dispelled.

Both Bruner and Dewey enunciated the contribution an organized body of knowledge makes to the problem-solving process, while at the same time pointing out that the application of the process should result in new knowledge. Both writers also asserted the creative nature of some aspects of the process. It was this attribute that received most attention in Bruner's comments, generally to the detriment of other features. Dewey's analysis appeared more balanced and inclusive. He actually denied the dichotomy of induction and deduction, asserting that the difference between them was merely one of direction, since oscillations between the two 'modes of thought' occur all the time as enquiry progresses (Dewey, 1938).

Later writers, sometimes acknowledging their indebtedness to Dewey, have attempted to utilize, clarify and expand the notion of reflective thinking. Nay (1971) found Dewey's five steps insufficient for describing what scientists do, so he increased the five to 14. After reviewing a number of professional books which discuss enquiry, Michaelis (1973) identified 13 separate skills in the enquiry process:

1. Recalling
2. Observing
3. Comparing/Contrasting
4. Classifying
5. Defining
6. Interpreting
7. Generalizing
8. Inferring
9. Predicting
10. Hypothesizing
11. Analyzing
12. Synthesizing
13. Evaluating

The Victorian primary schools Social Studies program, *Society in View* (Victoria, Education Department, 1974) saw a social enquiry as being made up of six major components: acquaintance; tentative explanation; testing; confirmation, rejection or modification; application; revision, extension or refinement. Under each component, sub-skills were indicated. These included questionnaire construction; listing, grouping and labelling information gathered from photographs and personal experience; and synthesis of ideas about a social situation.

It is evident that a bewildering array of terms are used to label the same cognitive processes. To add to the confusion, the same term is used occasionally by different authors to encompass different aspects of behaviour. Dewey has indicated five phases of behaviour which he labelled reflective thinking. Some researchers have continued to use this term (e.g. Massialas & Cox, 1966; Crabtree, 1967; Gross & Muessig, 1971). Others have coined new terms because they wish to give emphasis to different aspects of the activity. Bruner subdivided the process into two aspects by making a distinction between analytic and intuitive thought. Cleaves (1972) made a similar distinction in identifying what he called the creative component and the objective or scientific aspect. Cognitive tasks, thinking processes, problem-solving, human enquiry, productive thinking, creative thinking and scientific enquiry are all terms that have been used by different writers to label the enquiry process (Suchman, 1966; Taba, 1966; Covington, 1968; Allender, 1969; Kaltsounis, 1969; Herron, 1971; Covington, Crutchfield, Davies & Olton, 1972; Queensland, Education Department, 1972; Robinson, Tickle & Brison, 1972; Victoria, Education Department, 1974).

Critical Thinking and the Enquiry Process

While the attributes of reflective thinking and the enquiry process are generally considered to be identical, some disagreement exists as to what abilities are covered by the term 'critical thinking' (Henderson, 1972). Some writers equated it with reflective thinking (e.g. Massialas & Cox, 1966; Hunkins & Shapiro, 1967; Cordier, 1968; Blachford, 1973). Others saw critical thinking in a more limited context, suggesting it was an element in the total enquiry pattern (Ennis, 1962; Fenton, 1966; Orlandi, 1971). For Ennis critical thinking was primarily concerned with the analysis and assessment of data. In his discussion he listed 12 aspects:

(1) grasping the meaning of a statement;

(2) judging whether there is ambiguity in a line of reasoning;

(3) judging whether certain statements contradict each other;
(4) judging whether a conclusion follows necessarily;
(5) judging whether a statement is specific enough;
(6) judging whether a statement is actually the application of a certain principle;
(7) judging whether an observation statement is reliable;
(8) judging whether an inductive conclusion is warranted;
(9) judging whether a problem has been identified;
(10) judging whether something is an assumption;
(11) judging whether a definition is adequate; and
(12) judging whether a statement made by an alleged authority is acceptable.

(Ennis, 1962:83)

These components spell out more specifically the behaviours expected in phase four of Dewey's scheme. In general, critical thinking skills have been emphasized in those courses of study which rely strongly on interpretation of documents and concern students at the secondary school level (Fenton, 1966; Victoria, Education Department, 1968; Blachford, 1973).

A broader interpretation of critical thinking was given by Orlandi (1971). He included the identification of central issues and underlying assumptions, the evaluation of evidence and drawing of warranted conclusions, along with the ability to formulate reasonable hypotheses. These skills in combination with research skills (the location of information and the interpretation of graphic and symbolic data) constituted the set of problem-solving abilities.

Is There a Sequence to Enquiry?

Most writers who discussed the enquiry process saw dangers in presenting its attributes to students as a *set* procedure for solving problems. Dewey (1933) himself saw reflection as a chain of ideas moving towards a common end but cautioned that the sequence of the five phases of thought was not fixed:

In practice, two of them may telescope, some of them may be passed over hurriedly, and the burden of reaching a conclusion may fall mainly on a single phase, which will then require a seemingly disproportionate development.

(Dewey, 1933:116)

However, there appear to be different degrees of flexibility accepted by various authors. Klopfer (1971), for example, nominated four basic steps in the enquiry process: observing and measuring; seeing a problem and seeking

ways to solve it; interpreting data and forming generalizations; and building, testing and revising a theoretical model. These he saw as being taxonomical in nature. But within each step Klopfer detailed sub-categories of behaviour which did not necessarily appear in each investigation, nor always occur in the order listed. In this regard Klopfer's position was more flexible than that of Gagné (reported by Hills, 1970).

Gagné listed 11 levels of enquiry, commencing with observation and concluding with decision-making. Each level of behaviour subsumed all previous levels, while each level in turn was subsumed by all higher levels in the hierarchy of cognitive processes. Hills (1970:308), in reporting Gagné's work, suggested that the sequence of activities was crucial. For Hills, development could not be assured when the sequence was interrupted 'by skipping about in a helter-skelter fashion'.

By contrast, the teaching units of the Victorian Primary School Social Studies Program (Victoria, Education Department, 1974) indicated considerable flexibility in the application of the different components of a social enquiry. The gathering of information is a recurring process and the seeking of new approaches to group observations goes on almost interminably. Rather than being a neat linear process, as reflected in the logical exposition of attributes, the various categories of behaviour constantly recur as the enquiry proceeds.

How Appropriate are General Strategies of Enquiry?

Of even more moment than the question relating to sequence is the one concerned with the generality of the enquiry process described in Dewey's five phases and utilized by others with modifications. Is it, in fact, appropriate to generalize such a complex task as problem-solving into so few operations or aspects? Furthermore, is it safe to assume that these operations are applicable to any problem or issue?

From an examination of class units it is clear that both the Suchman (1964) enquiry training program and the Victorian Primary Schools Social Studies Program answer these questions in the affirmative. Curriculum developers in other school subjects have also utilized a general model of enquiry (Fenton, 1966; Victoria, Education Department, 1968; Klopfer, 1971; Cleaves, 1972; Bell, 1973). Admittedly, some are writing within the context of a single discipline, such as history, but others specifically record their belief that the skills they have identified are relevant to problem-solving

across a number of subject areas. Banks (1973) stated that the basic enquiry process is shared by all the social sciences, the critical factors which distinguish them are the questions posed and the content of the enquiry.

Because the model is flexible, Suchman (1964) argued, it can accommodate any problem. Yet a number of writers countered this by suggesting that only a superficial accommodation occurs because the categories of behaviour are so broad as to be almost meaningless and thus difficult to apply in practice (Schwab, 1964; Berlak, 1965; Connelly, 1969; Blachford, 1973). Using the analogy of games, Blachford pointed out that the ability to play games successfully means quite different things when spelt out in terms of tiddley winks and football. For Berlak, the procedures adopted by a clinical psychologist and a physicist in identifying a problem for investigation are different. He also noted the disagreement among many writers as to whether the objectivity in testing social science hypotheses is possible in testing hypotheses in history.

Some supporters of the general model, perhaps recognizing the lack of specificity, have expanded it beyond the original five phases and included many sub-categories of behaviour. Despite these modifications and additions, the critics of the general model are unlikely to be placated. With so many additions and so many different combinations of components being possible, the usefulness of such models becomes questionable. The very use of the term model may be inappropriate in such circumstances.

Those writers who emphasized the existence of different modes of thought argued that children should be inducted into the separate mysteries of each (Schwab, 1964; Berlak, 1965; Connelly, 1969; Hirst & Peters, 1970; Blachford, 1971; Herron, 1971, Robinson *et al.*, 1972). A general problem-solving strategy is not sufficient, although some authors would like it as well as specific enquiry modes. Hirst (1969), a prominent philosopher, argued that within certain domains of knowledge there are distinctive modes of reasoning. He suggested six: mathematics, physical sciences, human science and history, literature and the fine arts, morals, and religion and philosophy. Acknowledging some overlap, Hirst nevertheless maintained each involved elements which are irreducible to any others, singly or in combination. He claimed a moral judgement is not validated in the same way as a mathematical theorem, nor is the validation of an historical explanation attempted in the same way as a theological proposition. Hence there is a need, Hirst argued, to introduce students to all forms.

Blachford (1971) extended Hirst's classification to give geography a category of its own. He argued that geography is a distinct discipline, using a distinct mode of enquiry. Because geography is 'concerned' with spatial

problems, spatial classifications (mapping, regionalization) and systems', the investigator is obliged to use unique enquiry skills (Blachford, 1971: 219). Not all agree with Blachford, as he himself acknowledged. Simons (1971), writing in the context of geography, was more cautious. He saw the use of maps as the nearest thing geographers possessed to a special way of thought. Simons regarded the reading of maps as a specialized way of looking at evidence, an important aspect of enquiry. Maps are used by historians and geologists as tools of explanation, but for the geographer they may be a source of problems. Going a step further, Connelly (1969) noted there can be diversity of approaches to problem-solving within a discipline as well as between disciplines. This point is readily understood in social science programs where the issues being examined are in the context of disciplines such as sociology, politics and economics. However, Connelly believed different modes of enquiry are employed within disciplines such as biology and psychology. In the study of ecology he identified at least four classes of enquiries.

Berlak (1965) pointed out that the problems upon which the individual is expected to employ intellectual skills and abilities are extraordinarily diverse and complex. Knowledge about how people solve such problems is still extremely limited. In Berlak's opinion what is known is not sufficient to formulate adequately either specific or general enquiry strategies. However, he did provide a list of what he described as tentative and not necessarily mutually exclusive categories of problem types. These were personal, social, scientific, historical, practical-professional, and aesthetic. Note the similarity with the domains of knowledge given by Hirst. For each of these areas Berlak called for curriculum workers to develop 'context-specific' models of enquiry, each characterizing successful problem-solving behaviour. He used as an illustration the *Harvard Social Studies Project* where controversial issues in such areas as race, municipal politics and population control are used. In the development phase the writing team investigated adult discourse in the area of political controversy and the works of scholars in political science, law and philosophy. From this a model composed of several simple analytic distinctions was constructed.

An Ontario team (Robinson *et al.*, 1972) has gone ahead and identified six problem areas which, to their minds, employ different enquiry strategies. These are:

> *(a) Logical inquiry. The application of recognized rules of inference to given statements to produce universally acceptable conclusions. This approach typifies most of the mathematics that the student will encounter in high school and beyond.*

(b) Physical science experiments. Systematically intervening in a physical system to determine how it 'works'. This is the method of generating knowledge in most of the classical fields of science studied in high school and beyond as well as in such applied fields as engineering.

(c) Experiments involving the principle of random selection. Determining cause-effect relationships when the system in which we intervene may be permanently changed as a result of our intervention and where we cannot find an 'identical' control system. The most common application of this approach is in the newer behavioral sciences that try to unravel the causes of human behavior.

(d) Correlational analysis. Determining cause-effect relationships when we cannot undertake controlled experiments but can observe how the free variation of one factor is related to the variation in another factor. This is a common method of analysis in the social sciences (geography, history, sociology, anthropology).

(e) Case studies. Undertaking an analysis of the causes or results of individual events by amassing, categorizing, evaluating, and synthesizing large amounts of information relating to the event. In the traditional school program, this approach is used most commonly in history and geography.

(f) Real-life problems. Deciding which of two or more alternatives should be followed by analysing the consequences of each against a set of stated criteria. This model does not fit any specific school 'subject' with the possible exception of some guidance courses.

(Robinson, Tickle & Brison, 1972: 8)

Accompanying the specific strategies was a general enquiry model as the authors wished to emphasize the interdisciplinary nature of thinking. So students were introduced to a five-stage model, derived from Dewey, with problem identification, generation of alternative answers, collection and evaluation of information, synthesis, and drawing conclusions. Knowledge of this general procedure was seen to assist in recognizing the relationships existing between the six specific models. Because of such knowledge, it was expected students would see more readily the strengths and weaknesses of particular approaches.

Logical Sequence and Learning Sequence

As Smith (1957) indicated, Dewey's steps reflect a logical basis for developing in students the capacity to direct and control their own thinking. However, he warned educators that this was not necessarily a description of how thinking and learning actually occur. More recently Case (1974*a*) made the same point when discussing the inadequacies of the Gagné approach to learning. What then is the relationship between the developing ability of the child to think and the sequential steps of the enquiry process?

The work of Piaget and his co-workers has provided some data on this question. Piaget (1964), in mapping intellectual growth from childhood through adolescence, identified four orderly, sequential and qualitatively different stages; sensori-motor, pre-operations, concrete and formal operations. The latter, the most sophisticated, bears a not unexpected resemblance to the attributes of enquiry nominated by Dewey. In formal operations, thinking is marked by the formulation and testing of hypotheses which arise from the consideration of all the logical possibilities that could explain or provide a solution to a problem. Starting in the realm of abstract possibility, an individual at this level goes on to check his hypotheses with real evidence in a systematic manner. He is not bound to previous experience or reality; he manipulates ideas. Two writers specifically linked formal operations with the utilization of the enquiry process. Pledger (1972) did this in the context of controversial issues while Taba (1966) used Dewey's term 'reflective thought' interchangeably with Piaget's label 'formal thought' in discussing Social Studies problems.

The two earliest Piagetian stages bear little, if any, relationship to the steps of enquiry. The first stage, commencing with the birth of the child, features the co-ordination of action schemes. That is, the organization of spatial and causal relations and notion of object permanence. Later, in the second stage, the child's thinking is egocentric. He places great reliance upon the appearance of things. Within any one situation he tends to focus on one aspect to the exclusion of all others. When observing a sequence of events he centres upon the fixed states, paying no attention to the transformations occurring. Thus he offers contradictory explanations for an event and makes no attempt to resolve the conflict.

It is in the third stage, concrete operations, that some of the attributes of the enquiry process appear. The child in this stage is sensitive to the nature of the given problem and can give an answer free of contradictions. The answers offered are usually dependent upon the child's current or previous

experience. Thus, they may be seen as being superficial, and as concentrating on the surface features of a situation. Occasionally, however, answers are given which go beyond the concrete situation although they are limited in scope. They take the form of an extension of an actual situation, rather than an initial delineation of possibilities which are related to the given problem.

Many researchers have utilized the attributes of the Piagetian stages to rate responses to problems which could be described loosely as Social Studies-type situations (Lodwick, 1958; Case & Collinson, 1962; Goldman, 1965; McNaughton, 1965, Rhys, 1966; Hallam, 1967; McNally, 1970; Peel, 1971; Johnston, 1972; Whitehead, 1972; Jurd, 1973; Nettle, 1975). Some have specifically mentioned that their work confirms the sequence of stages proposed by Piaget. There are some investigators, however, who have expressed scepticism about the existence of stages discussed by Piaget. See Tanner (1960), Freyberg (1966) and more recently Brainerd (1973). A major difficulty is to determine how the existence of these stages can be confirmed empirically (or remain unconfirmed). Statements relevant to this issue have been made by Piaget (1968), Goldschmid (1971) and Beilin (1971).

The work of Piaget and others has not revealed whether the logically indentified attributes of the enquiry process correspond to the developmental sequence of thinking in children. The reports are not precise enough for that purpose. However, it does seem that some facets appear before others. The first three steps of enquiry are evident at the concrete stage of operations. Children at that level can identify the nature of a problem, generate possible answers and find some information to support the solution, although at a very elementary level. The suggested solutions arise from personal experience and not from an initial consideration of abstract possibilities. Nor is the full import of the problem always recognized. Only a few of the relationships existing in the problem are identified rather than the labyrinth of interactions. A systematic manipulation and evaluation of data is not present. These features of steps four and five in Dewey's enquiry paradigm appear at the formal operations level.

Ability of Students to Cope with Components of Enquiry

Piaget suggested that the period of formal operations commences at approximately 12 years of age, a time approximating the period of transfer from primary to secondary education in many national systems of education. Concrete operations ranges from about eight to 12 years thus covering the

majority of upper primary school children. However, with Social Studies-type problems, the advent of these stages appears much later. Hallam (1967) and Jurd (1973) placed the commencement of concrete operations at approximately 12 years. Peel (1971) and Rhys (1966) gave mean ages for subjects classified as concrete operators as 12+ and 11.8 respectively. For McNally (1970), 89 per cent of the grade six sample was rated at the concrete level. An identical rating was given to 67 per cent of Form II subjects. In a second and larger sample of Australian children McNally (1970) gained similar results. Whitehead (1972) found that the average level of performance at Form IV was at the transition point between concrete and formal operations. Jurd suggested 16.6 as the time of significant change in thought level. This is compatible with Hallam's finding of 16.2 and is only slightly higher than that found by both Johnston (1972) and Sinclair (1973). Goldman (1965), Peel (1971) and Rhys (1966) reported lower age levels for the onset of formal operations, but these were still two to three years later than the age levels suggested by Piaget.

Case & Collinson (1962) indicated formal operations could occur well before the age of 11 and cited cases to support their conclusion. Taba, Levine & Elzey (1964) appeared to support this general position when reporting that formal operations begin in a small way in grade two. Certainly, such conclusions are not supported by the research outlined above. Hallam attacked Case & Collinson on the grounds of inadequate rating of student responses. In challenging the classifications, Hallam insisted Case & Collinson were too lenient in their application of criteria for a formal operations rating. What constituted the small beginnings of formal operations noted by Taba *et al*. at the grade two level is not clear. Certainly there must be individuals operating at a particular cognitive level who are younger than the group average. That is, while the average age of formal operators might be 16 years, there may be cases of 12-year-olds also operating at the same level. However, they are relatively rare. McNally discovered a 10-year-old who answered at the formal level. Three cases out of 57 grade six children were reported by Whitehead as operating at the formal level on more than one test situation. Jurd, in that part of her sample drawn from grades five and six, noted subjects who had elements of formal operations present in their responses. Some children, for example, were able to go beyond the given data to support their conclusions but they were unable to co-ordinate all the appropriate behaviours in a total task and thus did not rate a formal operations classification. In an ingenious study, but not using the Piagetian framework of stages, Allender (1969) examined the ability of grades four,

five and six children to employ aspects of the enquiry process. Like Jurd, he reported that his subjects were able to complete components of the enquiry process. They were able to sense problems, formulate problems and engage in activities seeking information to make decisions.

In Piagetian terms, it seems highly unlikely that formal operations could be achieved by the majority of Australian children before the age of 11, particularly if the learning experiences currently available remain unchanged. With Social Studies-type materials such performance levels could not be expected until three to four years later. Thus, it is pretentious to expect present primary school children, indeed junior secondary students, to apply the enquiry process in its totality as described by Dewey.

Given the intellectual achievement levels of students it seems futile to introduce the differing enquiry procedures used in various domains of knowledge. Moreover in the present investigation, the time available to implement an enquiry program was very limited. The enquiry model adopted was therefore almost identical to that proposed by Dewey (1933) and more recently used by Robinson *et al.* (1972). Attention was given to problem-clarification, to the creation of alternative solutions, to the gathering of data, and to manipulation and evaluation of information to reach a decision. The actual learning experiences provided are detailed in Chapter 5. To develop these class activities, however, it was necessary to examine these questions:

> Why do children have difficulty in using the enquiry process?
> What are the constraints impinging upon the child's intellectual performance?
> Which, if any, of these constraints can be offset by the provision of learning experiences?
> What sort of school activities would be appropriate for that purpose?

The next chapter directs attention to these issues.

3. Cognitive Development and Training

To identify factors influencing cognitive performance, various curriculum projects have turned to different and divergent psychological viewpoints. Many curriculum writers have referred to the work of Piaget (Taba, Durkin, Fraenkel & McNaughton, 1971; Victoria, Education Department, 1974; New South Wales, Education Department, 1975). Piaget (1971) nominated four factors which account for intellectual growth: maturation of the nervous system, experience of objects and physical reality, social transmission, and equilibration. Generally, it has been left to others to identify the relevance of these factors for designing classroom experiences. In the present study it was the views of Pascual-Leone (1970) on cognitive growth that provided the theoretical base for the training programs.

Piaget and the Process of Equilibration

Piaget (1971*a*) claimed to be an interactionist. He refuted the notion that an individual's intellectual development is pre-programmed and that environment, experience and action are principally facilitating agencies, an interpretation of the Piagetian position proposed by Beilin (1971). For Piaget, intellectual growth results from the transformations of information derived from *both* genetic and experiential sources.

> *Cognitive structures consist neither of a simple copy of external objects nor of a mere unfolding of structures performed inside the subject, but rather involve a set of structures progressively constructed by continuous interaction between the subject and the external world.*
>
> (Piaget, 1970: 703)

The key to this process is the concept of equilibration. Equilibration is the process by which cognitive structures or schemes are altered. It is a self-regulating process in which the individual reacts to external disturbance. Commencing with a perceived discrepancy between what is seen and what is known, the individual acts to remove the conflict or disequilibrium through a reorganization of his cognitive structures. By doing so he achieves a new state of equilibrium or balance, a state where the intrusion of the environment becomes co-ordinated with his mental organization.

This process of adaption or reorganization involves two activities, one assimilation and the other accommodation. Both are involved in any intellectual act. A new situation is related to previous understandings of the problem, while simultaneously a new level of understanding is constructed to encompass this new circumstance. Assimilation is the activity which modifies or structures the environment to allow it to conform to the mental organization of the subject. Accommodation is the activity which modifies the mental organization of the subject in conformity with the demands of the environment. A person actively and spontaneously works to give meaning to his environment by assimilating the experiences into his current mental organization. Where a situation deviates from and is not sufficiently malleable to fit into his current mental structures, the latter will be transformed to accommodate the new situation.

Piaget (1970) pointed out that in cognitive adaptation there cannot be assimilation without accommodation and vice versa. While the ratio of assimilation to accommodation may vary, it is when there is more or less stable equilibrium between the two that a complete intellectual act exists. When assimilation outweighs accommodation, thought evolves in an egocentric direction. For example, to a child at play, a broom is a horse. Conversely, when accommodation prevails over assimilation, representations evolve in the direction of imitation. For example, a child reciting number names parrot fashion.

The functions of assimilation and accommodation not only concern certain specific tasks but also relate to the major developmental periods or stages of cognitive growth discussed in the previous chapter. The fundamental equilibrium between assimilation and accommodation is difficult to attain and maintain. Success depends upon the level of intellectual development and the nature of the problem encountered. Initially each stage is marked by considerable instability, but towards the end of the stage the organizational structures underpinning it have reached a relative state of equilibrium. Yet the equilibrium of the sensori-motor period is less stable than the equilibrium

of pre-operations, as is the equilibrium of pre-operations less stable than that of concrete operations. Thus, although relative equilibrium can be achieved within a stage, movement between stages occurs because the individual is automatically seeking a higher state of equilibrium. It is only when formal operations are attained that the most stable form of equilibrium is achieved. At that point the individual has a total, organized system which includes all possible transformations. Thus he can handle all environmental intrusions whether they are real or arise through his own thinking.

Other Piagetian Factors in Intellectual Growth

Piaget does not deny that heredity and maturation, the physical environment, and the social environment make a contribution to intellectual growth. Yet neither individually nor together are they sufficient to explain development. They must be in some mutual state of balance and they must be co-ordinated into a consistent and non-contradictory totality. Piaget (1970) maintained that this was the reason for appealing to a fourth factor, equilibrium, to explain intellectual growth.

For Piaget (1971*a*), maturation does no more than open up possibilities for the development of new structures formerly denied to the individual because of limitations in the nervous system. However, once the constraints have been removed the new freedoms have to be actualized through the action of the individual. The new reality still has to be constructed. This involves the intervention of other factors such as experience of the external physical environment and social interaction.

The contribution of physical experience is complex and always involves two aspects. Knowledge is derived from the objects actually manipulated and from the manipulations or actions exerted on them. The social environment, including the school environment, also makes its contribution. Within it, Piaget (1971*b*) emphasized the importance of the intellectual interaction between children during games and classroom activities. While acknowledging that pedagogical intervention can accelerate and complete spontaneous development, untimely intervention, Piaget (1970) cautioned, can at least temporarily cripple growth. He cited the example of parents teaching children to count before the children had any concept of number. He attacked the situation where knowledge is pre-sifted, and pre-digested by adults as though the process of transmission does not require a restructuring or assimilation by the child (Piaget, 1971*b*). To know is not simply to have a mental image of an object or event but to act upon it. Yet he did acknowledge

that adults do form part of the social environment which can stimulate and assist cognitive growth (Piaget, 1972). However, he did not spell out the features of that role. The physical and social environment can have an impact on development but only if the subject is capable of assimilating the experiences. He must have appropriate cognitive structures to which the experiences can be related and he must be aware of these physical and social intrusions if he is to operate upon them.

Relevance for the Design of Classroom Activities

The factors identified by Piaget as influencing intellectual growth lack sufficient detail to form the basis for the designing of day to day Social Studies activities in the classroom. Obviously, experiences need to be provided which relate to the current intellectual level of the child while at the same time challenging his present understandings. Primary school children and lower secondary students learn best through real situations and the actual handling and manipulation of objects and materials (Inhelder & Piaget, 1958). Activities, for example, which call for the manipulation of verbal propositions seem inappropriate for children still in the concrete stage of operations. Similarly, those activities which limit these same children to playing with objects without conceptualizing the operations being carried out are rejected. However, such advice is general rather than specific. There is great diversity in the real situations, objects and materials available to children. Are some more appropriate than others? Can children cope with all 'real situations' equally well? Should these activities be sequenced in some way? What should be done to ensure that the children see beyond the physical materials to the operations being performed? Piaget does not specifically answer these questions although occasionally his remarks have indicated his awareness of their existence (Piaget, 1971*b*). In the main, it is other writers who have extrapolated from his work to justify their solutions to these and other problems (e.g. Sigel, 1969; Athey & Rubadeau, 1970; Schwebel & Raph, 1973; Furth & Wacks, 1974). It should also be noted, however, that there are researchers who question the adequacy of Piaget's work for providing guidelines for designing classroom experiences (Sullivan, 1969; Aebli, 1970; Brainerd, 1973).

Piaget, (1971*b*) emphasized repeatedly the importance of the child *actively* manipulating his experiences. For Piaget, this meant the advocacy of discovery learning methods: 'each time one prematurely teaches a child

something he could have discovered himself that child is kept from inventing it and consequently from understanding it completely' (Piaget, 1970:715). Demonstrations or explanations by a teacher are an anathema (Piaget, 1971*b*). Overall Piaget denigrated the role of the teacher, almost implying that when the teacher becomes involved in the learning situation the child becomes passive and loses his creativity. This view appears to be based upon the assumption that a teacher's role can only be one of instruction. Piaget seems to have missed, for example, the organizational and facilitator roles that teachers might play in providing 'active' learning environments.

Memory

Piaget (1970) has noted a relationship between intellectual operations and memory. He saw progress in memory as being influenced by improvements in the operational schemes of intelligence rather than the reverse. However, as Sinclair (1971) pointed out, it was the organization of the information remembered that was determined by the particular cognitive structure of certain stages of intellect rather than the amount of information remembered. Others saw the amount of information remembered in the short term as a possible constraining factor on cognitive performance (e.g. Miller, 1956; McLaughlin, 1963; Posner, 1965; Flavell & Wohlwill, 1969; Pascual-Leone, 1970; Case, 1970; Bryant, 1971; Tomlinson, 1971; Halford, 1972).

Donaldson (1963) listed lack of concentration or a defect in immediate memory as a source of errors made by children aged nine to 13 when solving problems. Such children overlooked information previously deduced or given. Similarly, Dale (1970) observed that children, when attempting to solve Piaget's colourless chemicals problem, often forgot the combinations already tried. Other investigators have remarked upon the limited immediate memory capacity of their subjects (Lovell, 1961; Smedslund, 1966; Moray, 1967; Siegel, 1968; Case, 1970). Failure on any given problem arose because there was too much information for the child to co-ordinate or because the task had too many parts.

Miller (1956) maintained that a limited memory span imposes severe limitations on the amount of information that can be taken in, processed and remembered. Individuals can handle only a limited number of 'chunks of information'. Seven was the suggested maximum. Beyond that point input fails to be noticed. One chunk carries the same cognitive load as another although, as Miller explained, each chunk can be made up of a different

number of 'bits of information'; McLaughlin (1963) has taken up the same theme but has related memory span directly to the Piagetian stages. He postulated that each level can be simply defined by the number of different classes or chunks that can be distinguished simultaneously, and that this number of classes can be directly identified with a clearly measurable psychological characteristic — memory span. Each digit in a digit span test forms a chunk unless the subject is capable of mentally combining the digits into groups. At pre-operations a child must be able to handle two chunks simultaneously, at the concrete stage four chunks and at formal operations eight.

While accepting the criteria for pre-operations and concrete operations, Halford (1972) suggested that McLaughlin's requirement for formal operations was too demanding. Using such a criterion, Halford claimed formal operations would not be acquired until much later than 11 years in average children and never attained by adults whose memory spans did not develop beyond seven items. Putting an alternative proposal, Halford suggested that pre-operations require a memory span of two, concrete operations four, and formal operations six. The memory span must be equal to twice the number of dimensions relevant to the discriminations involved in solving a given task. Concrete operations were seen to be demanded in tasks of two dimensions and formal operations in tasks of three dimensions. Hence the need for memory spans of four and six respectively. While there is disagreement between Halford and McLaughlin on the precise relationship between size of memory space and stage of operations, there is agreement that memory space is an important factor in intellectual operations and that it can be measured.

While McLaughlin discussed at length the role of immediate memory in cognitive performance, he did not see it as the sole constraint. Although a necessary condition, it was not proposed as a sufficient condition. Yet he proceeded no further in identifying the other conditions. It was the work of Pascual-Leone (1970) that provided further insights.

The Pascual-Leone Model of Cognitive Growth

Pascual-Leone (1970) went beyond the constraints of immediate memory to discuss limitations imposed by the physiological capacity of the central processor, computing space, mental space or M-space. This is where the different chunks of information are co-ordinated, transformed and integ-

rated. However, the amount of information that can be attended to in a single act is limited. M-space is defined as the maximum number of activated schemes that can be co-ordinated at any one time. This capacity is assumed to grow in an all-or-none manner as a function of age in normal subjects. Thus, each successive developmental stage described by Piaget can be quantified in terms of, and is seen generally as a function of available M-space. The particular M-space values for the various stages have been derived from Piagetian data by what Pascual-Leone terms 'semantic-pragmatic analyses of tasks using symbolic logic' (Pascual-Leone & Smith, 1969). The M-space requirement for any given task is made up of the number of pieces of information necessary to generate the appropriate answer together with the subjects representation of the task instructions and his representation of the test situation as a whole. The latter two features Pascual-Leone saw as being constant across all tasks, although their precise M-space demands were unknown. The predicted values of M are set out in Table 1 (Pascual-Leone, 1970; Case, 1974c). The *a* in the table represents the task instructions and the subject's view of the problem situation. Case (1972*a*) called this the space required by the 'executive scheme'. The numerals represent the amount of information or additional schemes that can be co-ordinated at any one time.

TABLE 1
Predicted Values of M-space

Developmental Sub-Stage	Approximate Age	Maximum Value of M*
Early pre-operational	3 — 4	*a* + 1
Late pre-operational	5 — 6	*a* + 2
Early concrete	7 — 8	*a* + 3
Late concrete	9 — 10	*a* + 4
Early formal	11 — 12	*a* + 5
Middle formal	13 — 14	*a* + 6
Late formal	15 — 16	*a* + 7

The construct of M-space is not sufficient in itself to explain cognitive performance. There are two other aspects that Pascual-Leone proposed in his theoretical framework. The first relates to the units of information or schemes which the subject has in his repertoire, while the second concerns

*Pascual-Leone (1970) drawing attention to Miller's (1956) 'magical number seven' and the upper limit of M-space, suggested it could be a way of integrating Piagetian data with work on channel capacity of other experimental laboratories.

the subject's willingness and ability to utilize all the M-space available to him. If the subject does not have the appropriate set of reactions or schemes to apply to the given problem then he is not in a position to utilize the central processing space available to him. Alternatively, if the subject has the appropriate schemes and the required M-space, but is distracted by irrelevant perceptual features of the problem, then again, he will be unsuccessful.

Piaget (1970) distinguished between two types of knowing: figurative and operative. The same distinction was made in the Pascual-Leone framework*. Figurative schemes are those which represent reality as it appears without seeking to transform it. They are static internal representations of given things as these appear to the senses. They are mental images of items of information with which the individual is familiar or which he can recognize because of perceptual configurations. Case (1974*a*) saw them as being roughly equivalent to what Miller (1956) labelled chunks. Operative schemes are the active aspects of the mental structure through which reality is assimilated and transformed. They are the internal representations of higher order functions or rules which can be applied to one set of figurative schemes in order to generate a new set.

An important point is that *all* knowledge has an operational component. Knowledge is not attained by simply internalizing experience in the form of symbols which represent external events. Knowing is an act which transforms. Hence figurative schemes include an element of the operative about them (Furth, 1970). The two types of knowing are partial aspects of any real-life knowing, and never occur in isolation. The actual mechanics of knowing are not explained by either Piaget or Pascual-Leone.

In principle, a scheme has many degrees of complexity, from a single attribute to a complicated property list which is hierarchically organized. Yet a highly elaborated scheme takes up the same amount of M-space as a scheme with only one or two attributes. Case (1974*a*:546) observed that 'even though they (schemes) may not be unitary from the observer's point of view they may all be considered unitary from the subject's point of view'. This is an identical idea to that put forward by Miller (1956).

*Note that Pascual-Leone & Smith (1969) used the term 'scheme' to apply to both the figurative and operative aspects of cognitive functioning. They rejected 'schema' saying it suggested too strongly the idea of a template, a schematic form or image. Piaget distinguished 'scheme' from the term 'schema'. The latter term he uses with figurative knowing, the former with operative knowing. In this discussion the term 'scheme' is used for both the figurative and operative aspects of knowing.

Along with M-space and the repertoire of schemes held by the child, Pascual-Leone (1970) saw cognitive style as a possible constraint upon solution success. In particular he drew attention to Witkin's notions of field-dependence and field-independence. The salient perceptual cues in a problem may mislead some students into giving a response that is incorrect. Students receptive to such influences are known as field-dependent. Those who can resist the misleading perceptual characteristics or the wording of the problem, or both, thus making a logically correct response, are labelled field-independent. They have a tendency to activate the maximum number of schemes. The performance of children, on tests which measure these characteristics, appears to be remarkably stable over time and not amenable to training (Witkin, Dyk, Faterson, Goodenough & Karp, 1962; Witkin, Goodenough & Karp, 1967).

Educational Implications of the Neo-Piagetian Model

The Neo-Piagetian model of cognitive development suggests three factors which influence a child's performance on a problem task. Firstly, there is the information processing capacity or M-space a child has available. Secondly, there is the presence of a repertoire of schemes — executive, figurative and operative — appropriate to the problem. Thirdly, there is the existence of a cognitive strategy which encourages the child to ignore irrelevant but salient features in both the stimuli of and response demanded by the problem, whilst at the same time utilizing all available M-space. All three factors constitute serious constraints upon an individual's ability to apply the enquiry process as envisaged in the objectives of Social Studies teaching. If improvement in problem-solving is to be achieved, then these three groups of constraints must either be eliminated through the use of particular educational experiences or worked around by employing special cognitive techniques. Generally, Pascual-Leone had not related his views on cognitive development to classroom practice. This task has been taken up by Case.

Experience appears to have little influence upon the size of an individual's information processing capacity. M-space is largely hereditarily determined and grows as a function of age. The evidence, Case (1974*b*) argued, does not support the notion that the quantity of task-related experience to which the child is exposed will increase information co-ordinating capacity. Nor does the quality of general experience achieve this. In drawing these conclusions he cited his own earlier research, together with that of Parkinson (1969) and Gates and Taylor (1925). However, in a footnote, Case did concede that all

the variance in information processing capacity need not be tied to maturation. A substantial portion, he suggested, can be attributed to the interaction between maturation and other factors, such as 'general experience' and 'the natural processing tendencies of the growing child'. The view Case (1974b) has of field-independence is similar to his view of M-space. Seeing field-dependence as being relatively unmodifiable through experiences, specific or general, he linked its growth to maturational factors. Again, Case cited research findings to support his conclusions.

As field-dependence and cognitive space are relatively impervious to learning experiences, procedures must be identified which reduce their influence upon the problem-solving process. While the child's available M-space may be considerably smaller than the adult's, the child can be encouraged to use what he has more effectively. This can be done either by chunking information or segmenting the task. Both procedures reduce cognitive load and hence M-space requirements. It is known that the ability to chunk information increases with age (Flavell, 1971). Chunking occurs when discrete pieces of information are grouped together on the basis of common attributes to form a new, all-embracing single unit. Task segmentation involves breaking a problem down into a series of steps each requiring a smaller number of information units than pertain to the total task. The steps of the enquiry process, described in an earlier chapter, appear to lend themselves to the segmentation approach. However, because of the complexity of the tasks, segmentation within the enquiry steps would also be necessary if the operations required are to be brought within the capacity of the students. The conclusion drawn from one step of the task provides one unit of information for further processing at the next level.

The field-dependent student is not using all his available M-space. He accepts the answer which is perceptually appealing and simple in a situation where the complex response is the one that is wanted. To offset this tendency, three types of experience appear appropriate. The first involves training in strategies which help to reduce the sensitivity of the subject to particular elements of the stimuli. The second is the complement of the first: encouragement of the subject to consider all the given information and not just that which initially appears to be salient. Systematic scanning strategies therefore seem important. The third possible approach is to improve the student's figurative and operative schemes required to obtain a solution, thus reducing the cognitive demand of the task and so encouraging him to tackle it. It may then be possible to achieve a successful solution to a problem without utilizing all available M-space. While these three approaches can be

identified not all three may be necessary with all field-dependent students. Case & Globerson (1974) suggested different forms of intervention for students lacking the motivation to use all available M-space, as opposed to the student who reacts to limited aspects of the stimuli. Nevertheless, in a class situation where instruction is to groups not classified in terms of particular learning deficiencies, widely-based forms of experience are necessary.

The schemes component of the Neo-Piagetian model, unlike the M-space and field-dependence aspects, are considered to be amenable to change through the provision of appropriate classroom experiences. However, the composition of such experience was not detailed by Pascual-Leone; nor did he explain the mechanism of change. Case (1974*b*) suggested that practice is the key. Through it, old schemes will be modified or amalgamated to create new schemes. A number of schemes which are repeatedly activated together will be coalesced into one new, more sophisticated or inclusive unit. Basic to this action is differentiation and classification — the ability to pull apart the available information, noting similarities and differences, together with the ability to reconstruct it into fewer but more inclusive units.

Success in problem-solving is not only dependent upon the information stored in the student's memory and the ability to relate, transform and co-ordinate all this information but also by his ability to activate the relevant figurative and operative schemes. The student must be able to make appropriate discriminations and partitions. To do this, as Case (1974*b*) suggested, the student must have pertinent executive schemes or internal representations of a goal state. Put more simply, the student must have a plan to direct his field of attention: 'What am I attempting to do?' If the student has a clear understanding of what is expected of him, he can use his limited cognitive processing space more effectively. The search is more restricted and thus less demanding.

Kohnstamm (1970) argued that success in the Piagetian class-inclusion problem is influenced by the manner in which it is presented to young children. If examples of appropriate responses are given performance improves (Kohnstamm, 1963; Whitehead, 1972). Both Suchman (1961) and Duckworth (1973) suggested that children have been conditioned to act in a manner that places them at a disadvantage, in terms of performance, on creative, open-ended tasks. Duckworth believed class practices discourage children expounding their 'wonderful ideas'. Suchman noted a marked lack of autonomy in grade five students. Unless children are freed from their narrow perceptions of what constitute appropriate answers their perfor-

mance will not truly mirror their competence. Hence the importance of nominating suitable executive schemes for problem-solving*.

The Enquiry Process and the Pascual-Leone Model

Certain features appear to be essential in programs concerned with enhancing the ability of students to solve Social Studies-type problems. First, the program must provide an unequivocal view of what is expected of an individual in solving Social Studies problems. This provides the plan of action. It helps the child identify the figurative and operative schemes he must co-ordinate and transform. Not only does he know that he must reach a decision about a given problem but also he realizes he must consider alternative solutions, collect evidence and weigh it up. The paradigm of the enquiry model presented by Robinson *et al*. (1972) and discussed in Chapter 2 serves this purpose admirably.

Secondly, the program should present opportunities for development and enhancement of those figurative schemes associated with social issues. Most children are in a position to give an answer on many Social Studies problems, but the solutions they offer are superficial because of the restricted figurative schemes they have available. Without knowledge, solutions to problems cannot be found. Yet, in practice, children commencing school have had four to five years' experience of the real world so they already have a variety of schemes available. Generally, however, these schemes are very restricted, so that solutions offered to social problems are similarly scant (Kydd, 1969; Whitehead, 1971; Mugge, 1974; Nettle, 1975). If the student is expected to suggest a variety of solutions to a problem, then this will require him in part to have reasonably developed figurative schemes from which to define the nature of the problem and generate alternative solutions. To find solutions to the problem of 'poverty' requires an understanding of what 'poverty' is.

Consequently, the better the figurative scheme, the more successful the problem-solving performance is likely to be. Now this raises the question,

*It should be noted that Ausubel's notion of advanced organizers relates to this issue (Ausubel, 1960; Ausubel & Robinson, 1969; Allen 1970; Koran & Koran, 1973). The relevant subsuming concepts of the information to be learned are presented in advance. Acting as a form of pre-structuring, they direct attention to the figurative schemes the student will require if he is to get meaning from a verbal passage. They act as a bridge between what the student knows and what he needs to know. Ausubel's advanced organizers focus attention upon the content matter to be learned whereas the executive schemes detailed by Case are not as specific regarding the information component but highlight more the cognitive processes that will be required in a given problem.

'How can these schemes be developed and improved?' In this regard the Pascual-Leone framework on cognitive growth is no more helpful than the Piagetian. As was mentioned previously, the basic process appears to be the differentiation and classification of experience: the ability to look at old experiences, identify their attributes and relate them to new experiences. What is similar? What is different? From such an examination, one would hope that new relationships would be identified which strengthen the figurative scheme associated with that experience.

Sigel (1969) emphasized the importance of classification in cognitive growth. He believed children should be allowed to build classifications freely, to extend the quality and quantity of types employed. Drawing upon Sigel's conjectural views on classification Taba *et al*. (1971) advocated, in a Social Studies curriculum, the adoption of particular learning experiences. A number of Australian curriculum projects settled on a similar learning approach (Queensland, Education Department, 1972; Victoria, Education Department, 1974). In those programs the child was encouraged to isolate and enumerate pieces of information, to group these pieces according to common attributes, to label each group with a title that reflects its nature and then to write a sentence indicating the relationships existing between the groups. With such activity the child is simultaneously developing figurative and operative schemes.

Thirdly, provision in the program should be made to develop those skills needed in the process of reaching conclusions to identified problems. Without appropriate operative schemes the individual will be unable to engage in the enquiry process. Opportunity must therefore be provided for the individual to identify and practice suitable actions. Operative schemes associated with the generation of answers; the collection of evidence to support each suggestion and the procedures of information manipulation must be enhanced and developed.

In discussing constraints upon intellectual performances Pascual-Leone mentioned not only the schemes an individual holds but also the cognitive or channel capacity he has available to process and co-ordinate information. The M-space a child has is extremely limited, yet the issues he is concerned with in Social Studies can be highly complex. Usually they involve a great deal of information. One way around this dilemma is to chunk the information, to compress it into fewer and more inclusive categories so that it requires less channel space. Therefore the development described above in relation to enhancing figurative schemes seems appropriate. A second strategy to achieve the same purpose, the lightening of cognitive load, is to

break the task down into units of more manageable size, units which can be tackled in a sequence while being relatively independent of one another. Such segmentation seems apt with the enquiry process. The student considers the attributes of the problem situation, lists and groups information. From this chunked material he generates alternative solutions to the problem posed. Each alternative is considered in turn and evidence supporting or negating it is identified. The appropriateness of each alternative is examined in turn, in terms of the evidence gathered. And so the task of decision-making continues to the point where a final conclusion regarding the problem in made. As the enquiry progresses the student makes brief notes. By such means the cognitive load is again reduced. The details of a particular processing step are placed aside, freeing M-space, but those details can be recovered or referred to at any later step in the total process. Robinson *et al*., (1972), in presenting their students with a paradigm of the enquiry process, reinforced or developed suitable executive and figurative schemes. However, if students were permitted to record their answers on duplicated outlines of the paradigm, additional benefits in terms of reducing cognitive load would accrue. So fourthly, a program should utilize techniques which help to neutralize the disability of limited M-space suffered by students.

Fifthly, the constraints imposed upon problem-solving performance by field-dependence effects must be circumvented. In solving problems, some children direct their attention to perceptually salient but irrelevant features of the data. Other children are reluctant to make the intellectual effort required to consider all the information provided. The program must therefore show and encourage children to adopt procedures which overcome these constraining influences. Both the listing and grouping activity, and the task segmentation mentioned above, meet this requirement. The listing and grouping technique encourage the individual to scan the data systematically. Irrelevant but salient features will therefore tend to become less dominant. When the task is segmented into smaller steps, it appears to be and is, in fact, less onerous; thus the student is encouraged to utilize all the M-space he has available.

In terms of the Pascual-Leone model, programs which include the features detailed above could be expected to be successful in achieving improvement in both performance and competence of students when solving Social Studies problems. Therefore, these features provide the framework upon which the training programs used in the present research was built. The two experimental programs are described in Chapter 5. The next chapter examines previous research which has investigated the impact of enquiry training upon problem-solving performance.

4. Studies Investigating Training in Enquiry

Despite the inclusion of enquiry process objectives in Social Studies courses for more than a decade, relatively little substantial research has been undertaken to establish whether these objectives are realizable. With one exception, all the research that has been undertaken has been done overseas.

Rationale Underlying Training Studies

Few researchers have provided explicit statements about the psychological or learning model from which their training programs were derived. Some adopted an eclectic approach in drawing from a number of different and diverse views on what constitutes appropriate learning experiences (for example Taba, 1966). Others were more pragmatic in that they included in the final program any activity or procedure which appeared to work in classroom trial sessions (Jenkinson & Lampard, 1959; Mason, 1963; Anderson, 1965). A slightly more systematic approach was adopted by Suchman (1961) where he identified common weaknesses in the ability of children to use the enquiry process. He then developed a program which aimed at overcoming these deficiencies. Others based their programs upon the one developed by Suchman (Butts & Jones, 1966; Scott, 1973). Robinson *et al.* (1972), when justifying the presentation of an overview of the enquiry process to students prior to giving instruction in the separate skills, did mention Ausubel and his views upon the place of organizing ideas. However, Robinson went no further in relating Ausubel's cognitive theory to the enquiry training program. Cousins (1963) based his program on a logical model of enquiry without any overt consideration of how children learn.

In terms of such things as teacher involvement, materials used, attributes of the enquiry process emphasized, and duration of training, the different investigations showed considerable variation.

Teacher Involvement

With a number of studies, teacher participation in class activities was minimal. Ripple & Dacey (1967), Wardrop *et al.* (1969), Treffinger & Ripple (1970) and Gray (1972) all used what could be loosely labelled as independent student assignments. Gray used a branching program designed around multiple choice questions. The Wardrop investigation made use of an early version of the Berkeley *Productive Thinking Program*. There the children worked from booklets produced in comic style. The problems to be investigated, together with appropriate information were given within the context of discussions between two children and their uncle. A modified version of this program was used by Ripple & Dacey. The teachers only distributed the lesson material and answered procedural questions. Class discussions were not held. Anderson (1965) used the techniques of programmed instruction in the form of a script read by the teacher to one child at a time. This teacher participation arose because the subjects were only in grade one. Although Massialas & Zevin (1964) and Blatt & Kohlberg (1974, reported by Rest) did not use programmed instruction, teacher involvement was relatively minor. The training technique used by Blatt & Kohlberg relied upon discussion between students of differing ability levels. Through such interaction intellectual progress was expected. The teacher's role in the Massialas & Zevin work was mainly non-directive in nature, being concerned with such things as re-phrasing questions and recognizing students.

There have been investigations which, in relation to the nature of teacher involvement, contrast with those noted above. In the studies of Taba (1966), Hunkins & Shapiro (1967), Olton & Crutchfield (1969) and Wallen *et al.* (1969), teachers were intimately involved in the training sequence. An essential part of the Taba program was the use of particular types of teacher questions during the learning activities. Because of this, teachers using the program were given special training before introducing it in the classroom. A similar situation pertained in the Wallen investigation as this was an extension of the Taba work. While teaching strategies were not emphasized to the same extent in other investigations, teachers were expected to stimulate and guide discussion in class sessions (Jenkinson & Lampard, 1959; Cousins, 1963; Massialas, 1963; Hunkins & Shapiro, 1967; Olton & Crutch-

field, 1969). In the Suchman (1961) program a tape recording of the previous session was played to the class and the teacher evaluated the success of the work. Attention was given to the nature and appropriateness of the questions asked by the children and emphasis was given to the elements of the enquiry being employed. Crabtree (1967) was specifically examining the influence of the teacher communication and the pre-arrangement of the learning environment upon the productiveness of children's thinking. Robinson (1973) saw the teacher as playing an active role in the training sessions although he recognized the role as being either a 'laying out of content' or a 'drawing it out' from students.

Enquiry Skills Taught

Various enquiry skills have been stressed in the different programs. Some projects have been concerned with the full range of skills, including the generation of answers, the seeking of information, the weighing up of evidence and the drawing of conclusions (Massialas & Zevin, 1964; Taba, 1966; Covington *et al*., 1972; Robinson *et al*., 1972). Both Shulman (1965) and Allender (1969), however, correctly asserted that most training studies have ignored the initial phase identified by Dewey, namely problem-identification. The parameters of the situations in which problems were embedded were defined by teachers and the attention of students was directed towards particular issues. In addition, it is evident that certain skills receive more attention, at least in terms of time allocation, than others. No doubt, in many cases this time allocation arose from observations made during the development of materials to the effect that children found some skills more difficult than others. Suchman (1961), for example, stressed question-asking. However, underlying this particular skill are others, such as observing and thinking of possible explanations to the phenomenon observed. For Taba (1966), the ability to classify was crucial in cognitive performance, so listing, grouping and labelling abilities were basic to her program. The valuing of particular sets of skills has been also reflected in the evaluation instruments used to measure success of the training episodes.

There are other studies which had more restricted training objectives. Higgins (1974) was concerned with improving the ability of children to draw inferences from given data. Arising from such training, children were expected to create more answers to given problems. Ojemann & Campbell (1974) were also concerned with teaching children to find alternative solutions to given problems. But, in addition, they wanted children to examine the alternatives in terms of their consequences in common social situations.

Anderson (1965) was teaching his subjects the technique of successively varying each factor in a problem while holding all other factors constant. A similar purpose was held by both Bredderman (1973) and Case & Fry (1974). Gray (1972) trained children to manipulate data. Although Gray saw this as a limited objective, it did involve the subjects in the production and evaluation of possible answers to problems. Hence Gray's program was as comprehensive as some of the programs which claimed to be involved with the total enquiry process.

Content and Training Approaches

Training approaches also differed between studies. For example, the authors of the Berkeley materials stressed that their program considered the enquiry process as a totality (Covington, n.d.). The separate elements of the process were taught within the context of particular problems. They were not taught first in isolation. A similar situation applied to Taba (1966). By contrast, Case & Fry (1974) initially taught subjects to suggest counter-explanations to a given set of facts. Later, students were instructed in designing experiments for which no counter-explanations were possible once the facts were obtained. In the final sessions these two skills were brought together. With the Ontario materials the students were given an overview of the total process and then separate sessions were devoted to the different components required to build up performance on the total model (Robinson *et al.*, 1972; Robinson, 1973). For example, when considering real-life problems, single sessions were devoted to asking more general questions, to brain-storming and to using common sense in eliminating alternative solutions. Each of these elements was constantly referred back to and related to the enquiry model as the lessons progressed. Both approaches, the element and total strategies, appear to have been successful in achieving improvement in student performance, as the results discussed below indicate.

There were also differences between the studies in the content areas of the problems investigated and the nature of the stimulus material within which the problems were embedded. Suchman (1961), Mason (1963), Butts & Jones (1966), Scott (1973), and Bredderman (1973) worked within the physical sciences. Others can be loosely grouped in the social science category (Cousins, 1963; Elsmere, 1963; Massialas, 1963; Taba, 1966; Hunkins & Shapiro, 1967; Wallen *et al.*, 1969; Higgins, 1974). Taba and Wallen *et al.* drew from a number of the social sciences including sociology and history. Hunkins & Shapiro dealt with issues such as freedom of speech, freedom of worship, privacy and equality. Ojemann & Campbell (1974) and Blatt & Kohlberg (1974) used moral dilemmas while Robinson *et al.* (1972)

drew from a number of areas including logic, the social sciences and the physical sciences. Case & Fry (1974) in their five units drew from physics, psychology, biology, engineering and sociology. The Berkeley materials originally centred on 'crime'. In more recent versions of the *Productive Thinking Program* this has been extended into other fields (Covington *et al*., 1972). Massialas & Zevin (1967) developed activities in the fields of English, geography, history, physics, music and criminology.

Diversity also existed between studies regarding the form of the data. Written materials seem to predominate although there were notable exceptions. For Suchman (1961) and Butts & Jones (1966) movie film clips provided the source of the problem. Line drawings with comic style balloons of conversation provided the basic material in the Berkeley program. Others such as Cousins (1963), Elsmere (1963), Massialas (1963), Hunkins & Shapiro (1967), Gray (1972), Robinson *et al*. (1972), and Ojemann & Campbell (1974) relied upon written materials. The Taba (1966) and Wallen *et al*. (1969) studies used the *Taba Social Studies Curriculum*. This course listed library reference material, and, less often, filmstrips and motion pictures that could be used (Durkin, Fraenkel & Tanabe, 1969). The Brederman (1973) investigation was one of the few studies where the problem was presented in the context of physical apparatus. To obtain an answer the child had to manipulate the apparatus.

Duration and Intensity of Training

The hours of instruction given in the various programs differed markedly. Some gave intensive training for a short period of time while others used regular sessions spread over the school year. Wallen *et al*. (1969), for example, reported that over the year, three of the experimental teachers devoted an average of four hours per week, five averaged three hours per week, one averaged two hours and another teacher one and a half hours for the week. A school year was also the time duration of the investigations carried out by Taba (1966) and Robinson *et al*. (1972). Weekly time commitment was not given in either case. Subjects in the Scott (1973) investigation engaged in an enquiry training program over three years. In science classes throughout each school year the subjects received approximately one hundred minutes of training per week. The Massialas program (1963) was spread over 18 weeks.

By contrast Gray (1972) had only five training sessions, each of one period length. The distribution of these sessions was not indicated. Butts & Jones (1966) held training sessions for each school day for three weeks. A

similar period of time was used by Hunkins & Shapiro (1967), 16 class days. Wardrop *et al*. (1969) had one lesson per day for four weeks.

For the achievement of such a complex set of abilities encompassed by the label 'enquiry' the time devoted to training by these latter studies seems inadequate. Yet as the discussion of results shows, almost all claim to have achieved improvement in performance. In fact, Ripple & Dacey (1967) commented upon an earlier finding of Crutchfield & Covington where significant differences between instructional and non-instructional groups on criterion tests were found after lessons four and 10. Why these improvements should have occurred was not always clear. They could have been due to teacher enthusiasm, teacher bias or a Hawthorne effect. These possibilities are explored later in this chapter.

Results

While it does emerge that pupil performance on aspects of the enquiry process can be improved through training, only limited patterns of consistency are evident across studies. In part, the difficulties in interpreting results arise from the absence of a detailed rationale underlying the training program, lack of detail concerning the nature of the training provided, insufficient attention to experimental design and the limited nature of the evaluation instruments used.

Treatment Effects

The improvement in student problem-solving behaviour, in almost all investigations, was associated with the training provided. One exception was that of Mason (1963), where success at one grade level was attributed to teacher effects. However, the training program appears to have been effective at the grade five level in teaching 'critical thinking'. Hunkins & Shapiro (1967) concluded that elementary school children can be taught to think critically. In this instance critical thinking referred to the ability of students to analyse a situation and be consistent in their actions in relation to their choices. Scott (1970) suggested style of categorization was influenced by enquiry training which emphasized an analytical approach to problems. Attention to detail became more acute. Verbal fluency and flexibility were increased. Success was also achieved by Butts & Jones (1966) in enhancing problem-solving abilities, although their success appears broader in scope because of their more comprehensive test instrument.

Taba (1966) reported that those children trained to discriminate, to infer from data and to apply known principles to new situations, were superior to those untrained. However, there were inconsistencies in the results which she suggested could have been influenced by the inadequacies of the tests, the imbalance in the composition of the sample groups or variations in teaching style over and above the variable of training. Wallen *et al*. (1969) found differences between the control and treatment groups on some measures. The treatment group showed greater gains from pre-test to post-test than the control group in the ability to interpret data and make legitimate inferences. Higgins (1974) also found that inference-making behaviour of nine- and ten-year-old children could be significantly improved. However, significant differences between groups were not found by Wallen on the ability to use, in new situations, the generalizations emphasized in the Taba curriculum, nor to explain given phenomena or events after identifying the relevant elements or relationships.

Training effects were also found in three investigations examining the adequacy of the *Productive Thinking Program* in achieving its given objectives (Olton & Crutchfield, 1969; Piper, 1969, cited in Covington, n.d.; Wardrop *et al*., 1969). Wardrop *et al*. noted the superiority of the treatment group on 30 out of 40 measures included in the study. Of these, 13 proved to be statistically significant, with 11 in favour of the treatment group. Superior achievement by the treatment group was noted on measures concerned with problem-solution, number and quality of ideas produced, intellectual persistence and sensitivity to discrepant or puzzling facts. These results were described by one member of the project team as modest (Covington, n.d.). In the second major study by the Berkeley group, the training outcomes were more substantial (Olton & Crutchfield, 1969). Reporting on a sub-sample of 50 subjects from a larger study, Olton & Crutchfield noted three specific results from the training program. First, the treatment group was better able to account for puzzling facts. Secondly, they were able to grasp the significance of a given clue to eliminate alternative suggestions to solve a hypothetical archaeology problem. Finally, they were able to suggest more causes of poverty and how it might be ended; than was the control group. Both groups, however, were roughly equivalent in basic knowledge about poverty. In addition to these results on the elements of the enquiry process, a composite score was calculated. This the authors saw as a very comprehensive measure of the thinking performance, reflecting achievement on 23 different indices of productive thinking. The differences between the treatment and the control groups on this combined measure was significant at the .01 level.

Piper (1969), essentially replicating the Olton & Crutchfield study, also found significant differences on a composite score, calculated on the basis of 14 variables, between performance by treatment and control groups. Using a modified *Productive Thinking Program*, Ripple & Dacey (1967) found a significant difference between instructed groups and non-instructed groups in time taken to reach solutions to problems. But there were no significant differences on measures of fluency, flexibility, imagination or originality. A more recent study involving Ripple (Treffinger & Ripple, 1970) concluded that the instructional materials did not influence pupil's verbal creativity scores to any appreciable extent. With respect to other problem-solving criteria there was little evidence for the effectiveness of the instructional materials as used in that study. Moreover, there was no support for the assumption of positive transfer from the program to problem-solving in another subject area; in this case arithmetic. Obviously these findings conflict with some of those of the Wardrop *et al*. and the Olton & Crutchfield investigations.

Robinson *et al*. (1972) report that evaluation of their program has been undertaken over a number of years but present scant information as to its success. The authors do state, however, that the program has had a large effect upon the ability of students to recognize relationships and the implications of these relationships, to identify limitations in a given set of experimental controls, to use multiple criteria in decision-making, and to recognize the restrictions involved in applying the conclusions formed from one situation to another.

Both Anderson (1965) and Case & Fry (1974) found it possible to train subjects to control variables in a problem task. The differences between treatment and control groups in both studies were highly significant. Although examining the same ability, Bredderman (1973) found no significant differences between groups.

Two other investigations claim success in training children to handle moral dilemmas (Blatt & Kohlberg, 1974; Ojemann & Campbell, 1974). Ojemann & Campbell concluded from their results that planned learning experiences can make a significant difference in the development of factors important in making moral judgements. It will be recalled that they were particularly interested in encouraging subjects to think about alternative solutions and the consequence of such decisions.

In summary, it seems that certain attributes of the enquiry process are susceptible to training. Yet it is still unclear whether performance improvement on the elements necessarily relates to improvement across a total task,

and whether improvement in some of these elements is more crucial than others in gaining overall growth. Indeed, as will be seen below, it cannot always be said with confidence that factors other than the training program were not influencing performance.

Other Factors Influencing Performance

Several studies have commented upon the relationship between the ability of students to solve problems and variables such as intelligence or mental age, vocabulary scores, sex, and classroom environments (Harootunian & Tate, 1960; Dale, 1970; Peel, 1971; Johnston, 1972; Whitehead, 1972). Also, many of the training studies in pre- and post-testing phases have noted such relationships. Wardrop *et al.* (1969) found from pre-testing sessions that students drawn from what were described as 'facilitative environments' performed better on tests measuring number and quality of ideas produced to given problems, compared to those from environments described as 'non-facilitative'. However, only one similar effect was found out of 40 post-test measures. Mainly, Wardrop argued, this was because the treatment was particularly effective with students in non-facilitative environments. Significant interaction between environment and treatment supported this argument. Crabtree (1967) found that a higher degree of structuring in the learning situation resulted in a significant effect upon measures of convergent thinking, while lower levels of teacher structuring elicited more divergent responses.

Significant sex differences were also identified on pre- and post-measures of the Wardrop study. With the exception of three test measures which all related to the same problem, the performance of girls was superior to that of boys. This finding, Wardrop *et al.* felt, was consistent with other findings regarding performance of females versus males on verbal ability tests. The same study reported that the general level of intellectual ability of students correlated significantly with problem-solving performance.

It seems likely that certain characteristics of subjects and the learning environments from which they are drawn and the experimental situation in which they are placed can influence performance on problem-solving tasks. Because of this, it is important to establish it these same factors or others can influence the success of a training program. Should they do so, then in further experimental studies such factors should be controlled so that a less contaminated measure of treatment effects can be gauged. Preliminary leads have been given in some studies.

Both Butts & Jones (1966) and Wardrop *et al.* (1969) found no significant interactions between IQ measures and treatments. Wardrop *et al.* pointed out

that, in contrast to the presence of IQ effects on test performance, no consistent pattern of interactions between IQ and treatments existed. While a similar conclusion was drawn by Butts & Jones, they were more cautious, believing their results may have been influenced by the size and homogeneous nature of their sample. For them the issue was not closed. Yet Wardrop *et al.* specifically commented upon the representativeness of their sample, covering the entire spread of ability in a total of 44 classrooms. Earlier Suchman (1961) had noted that not all children benefited equally from enquiry training. However while anticipating that children who were 'intellectually gifted' would get most from the program, he found they were often the least willing to do so. Such children were often seeking an immediate solution. This latter conclusion of Suchman conflicts with that reported by Massialas & Zevin (1967), that on one problem the low IQ group was more creative in its response but did not persist in following through its hunches. The high IQ group had fewer and less original explanations but was more persistent in systematically following its explanations through to a logical conclusion. After blocking levels of ability into two groups Olton & Crutchfield (1969) found both groups benefited from instruction although the high ability level slightly more so. The statistical significance of these differences was not presented.

Only Mason (1963) reported a teacher effect. He inferred that the difference between his control and treatment groups at the fourth grade level was due to a teacher factor. The treatment group, utilizing teachers who had developed the training materials, performed better than either the other treatment group or the control group. While no other study specifically reported upon teacher effects, Covington (n.d.), when reviewing research on the Berkeley materials, highlighted the difference in the magnitude of results between the investigations of Wardrop *et al.* and Olton & Crutchfield. This difference, Covington suggested, could be due to teacher involvement, since in the first-named investigation the training materials were used independently of the class teacher. With the second study the teacher was closely involved. Caution, however, should be used in interpreting this relationship, as the duration of training sessions and materials used also varied between studies.

The findings of Mason (1963) indicated that some grade levels benefited more than others from training. Ripple & Dacey (1967) also suggested that instructional materials may become less effective in developing creative thinking abilities as grade level increases. Such a possibility gained no support from Treffinger & Ripple (1970).

Nature and Appropriateness of Evaluation Instruments

Few training studies which examined treatment effects used measuring instruments which attempted to evaluate the student's performance on the problem-solving task as a totality. Most studies used tests which examined elements of the process. Examples of such abilities are the generation of solutions, the achievement of correct solutions and the identification of relevant evidence. The circumscribed evaluation instruments pertained not only to those investigations with limited training objectives but also to those which set out to enhance the total enquiry process (Butts & Jones 1966; Taba, 1966; Wallen, *et al.*, 1969; Robinson, *et al.*, 1972). The studies associated with the Berkeley *Productive Thinking Program* used an impressive and highly creative array of tests, yet each test was limited in the enquiry objectives it was testing (see Wardrop *et al.*, 1969). For example, from an account of the migratory behaviour of a hypothetical flock of birds, the student had to identify puzzling facts in the story. Later, after having some of these facts pointed out to him, he had to devise explanations to account for them. The student responses were scored separately on each variable. Covington (1968) and Olton & Crutchfield (1969) did go on to calculate an overall problem score but this was derived from a pooling of scores on each of the variables measured on the various tests. Summing of scores across measures of specific abilities to gain a single general enquiry score was a technique adopted by a number of investigators (Shulman, 1965; Ripple & Dacey, 1967; Treffinger & Ripple, 1970). Such a procedure demanded the use of reliable instruments which measured different aspects of the enquiry process. Even if these requirements were met, the question concerning the relationship between the summed performance across separate tasks and performance on global problem-solving measures would need to be answered. Does the sum of the parts equal the whole? Ripple & Dacey used a battery of creativity tests on which the subjects were rated on flexibility, originality, fluency and imagination. They also used Maier's Two-String Problem to measure problem-solving performance but here attention was directed to the achievement of the correct solution and the time taken to gain it, rather than to the thought processes used.

The two criterion tests used in the Taba (1966) study related to the ability to interpret data and apply principles to new phenomena. Associated with the ability to interpret data were skills of inference-making and the judgment of the validity of inferences made. Separate scores were calculated on each aspect. The ability to apply principles to new phenomena was included because the author believed generalizations made in the social sciences are

often inclusive and stereotyped. Part of the training program was to encourage children to recognize the limits of data and to refrain from over-generalizing, or conversely, from being over-cautious. The second test, *Application of Principles Test*, used concepts and generalizations which the student had developed through the learning program. It presented new situations and asked the child to select the generalization which explained the occurrence or event portrayed. Similar limitations in the objectives of enquiry assessed by other investigations can be documented.

The objectives being tested in most studies reflected a value weighting placed upon some abilities in the total enquiry process by the investigators. Taba and the investigations associated with the *Taba Social Studies Curriculum* stressed the importance of categorization in intellectual growth so it was reflected in the instruments used to evaluate success of the program. As the name of the *Productive Thinking Program* implies, the Berkeley studies accentuated the creative component in enquiry (Covington, 1968; Olton & Crutchfield, 1969; Wardrop *et al.*, 1969). The tests used, in addition to rating explanations given as the consequences of problem-solutions, examined the number and quality of ideas advanced by subjects to solve problems and the ability of subjects to re-organize the elements of a problem, all of which are creative tasks. For Suchman (1961), with his concern for developing autonomous thinkers, the evaluation centred upon the questions asked; the fluency, control, precision and autonomy in finding answers to filmed demonstrations.

Aside from the objectives being evaluated the measurement techniques used in some studies appear to place further constraints upon the interpretation of results. A number of studies used multiple choice-type tests (Jenkinson & Lampard, 1959; Cousins, 1963; Cox, 1963; Massialas, 1963; Nelson & Mason, 1963; Taba, 1966; Hunkins & Shapiro, 1967; Ojemann, & Campbell, 1974). The multiple choice technique significantly changes the nature of the mental operation carried out by the subject, compared to situations examined in Social Studies curricula. While it is true that items presented in multiple choice form can be testing more than the recall of factual knowledge, more constraints in the nature of an acceptable response are inevitable. In addition, such tests appear to place a premium upon convergent thinking. This constraint seems inappropriate where interest is in the process of mental operations rather than a particular product. This is not to say that multiple choice instruments have no place in investigations into problem-solving but, due to their inherent limitations, such tests need to be supplemented by other more "open' evaluation techniques. The inadequacy

of enquiry measures in physical science was noted by Nelson & Abraham (1973). They proceeded to develop their own open-ended instrument which provided separate scores on students' ability to gather data, to project into an unexplored situation from an explored field, to test the validity of an inference and to form groups having some common specified observed property.

There are studies which have employed more open measures. Gray (1972) used open-ended test situations based upon the comprehension-type passages advocated by Peel (1966) and used by other investigators inspired by Peel's work. A similar technique was used earlier by Elsmere (1963). Butts & Jones (1966) and Suchman (1961) used open-ended situations in the physical sciences, as did a number of studies using the Berkeley program (Oltoh & Crutchfield, 1969; Treffinger & Ripple, 1970). Transcripts of class lessons were used by Cousins (1963), Cox (1963), Elsmere (1963) and Massialas (1963) to rate the frequency with which elements of the enquiry process occurred.

Reliability of Instruments

It has been noted above that some studies employed multiple choice items while others used open-ended situations where scores were based upon single criteria. There are obvious advantages in the first approach in terms of reliability of marking but the appropriateness of such measures, even when summed, in terms of the total problem-solving process, is questionable. This issue is taken up again in a later discussion on scoring student responses to problem tasks (see Chapter 6). However, in addition to this difficulty, a number of investigators have indicated doubts about the instruments they employed. Taba (1966) pointed out that her criterion measures had certain deficiencies in respect to sensitivity and stability. Jenkinson & Lampard (1959) discovered that the ceiling of the STEP tests, the criterion measure of the investigation, was too low. Many of the subjects at the grade six level identified the correct responses on almost all items at the initial administration of the tests. With the Treffinger & Ripple (1970) investigation a number of deficiencies in the measuring instruments were noted. While claiming the *Torrance Tests of Creative Thinking* were 'the most practically useful instruments available, at the time, to measure fluency, flexibility and originality', the authors acknowledged that the tests did not 'comprehensively assess creativity and that these tests are accompanied by a number of technical, procedural difficulties'. Treffinger & Ripple also noted that the

Kuder-Richardson Formula 20 reliability coefficient for the Arithmetic Puzzles, a transfer task, was lower than desirable. With the general problem-solving tasks there was a high level of difficulty. This difficulty level, Treffinger & Ripple argued, may have had the effect of masking differences between treatment groups.

Transfer and Durability of Training

To determine the real success of training attention needs to be given to factors beyond the scores obtained on post-test measures. Piaget (1964) detailed three important criteria. The first is durability of the ideas taught. The second concerns the extent to which the ideas are generalized by the child, and hence, transferable to new situations. The third criterion is the spontaneous level of operation before the learning experience was introduced. Durability and transfer as necessary criteria have also been raised by other investigators (Bracht & Glass, 1968; Inhelder & Sinclair, 1969; Snow, 1974).

All three criteria were not always met in the training studies reviewed in this chapter. Generally all subjects were pre-tested. Attention was given to durability by Covington (1968), Olton & Crutchfield (1969), Bredderman (1973) Scott (1973) and Blatt & Kohlberg (1974). Olton & Crutchfield found that the gains in thinking skills produced by eight weeks of instruction evidenced in the performance of the treatment group was still evident on 10 variables more than six weeks after instruction had ended. However, Covington (n.d.) indicated that the margin of superiority favouring the instructed students had tended to decrease over time. After one month Bredderman found that the slight supremacy of the treatment groups over the control had dissipated. Scott's training program concluded in 1965 but he found that significant differences between experimental and control groups had persisted on one sub-score of the Sigel *Cognitive Skill Test* on retesting in 1966 and 1971 (Scott, 1973). The particular skill concerned the student's analytical behaviour. Scott satisfied himself by questioning students that further enquiry experience had not been gained either for the experimental or control groups in the intervening period between training and delayed testing. The Blatt & Kohlberg (1974) investigation found that students who had advanced after 12 weeks of training remained advanced 12 weeks later.

Regarding the transfer effects, few studies examined whether specific skills which had been taught carry across into new situations or other subject

areas. Robinson *et al.* (1972) identified three levels or dimensions of evaluation*. First, there are measures of take. These concern the extent to which the student has learned what he has been specifically taught. Many of the tests used by Olton & Crutchfield (1969) and Wardrop *et al.* (1969) fall into this category. Secondly, there are those situations which involve transfer within a range of problems. A new situation is presented to the subject which requires the employment of strategies and skills developed in the learning program. Since the Bredderman (1973) study involved training the subjects to control variables in physical science situations the use of a social science problem where control of variables was necessary to gain the correct solution could have been used as a measure of transfer. This was not done. Case & Fry (1974) actually used both social science and physical science tests but gave experience on both types of situations in training sessions so the opportunity to measure the quality of transfer was lost. Covington (n.d.) considered the Bird Migration Problem used in a number of the Berkeley-inspired investigations to be a transfer task. This was because the *Productive Thinking Program* contained no material of the migratory behaviour of birds or any related concepts. So in this case the student had to apply thinking skills learned in one area to another with which he had little or no previous experience.

The third and final dimension of evaluation identified by Robinson *et al.* (1972) concerns tasks which give no clues that particular strategies taught in class should be used. Such tasks involve a much higher level of transfer. Ideally, as Robinson *et al.* pointed out, these tasks should not 'smack of school'. Ripple & Dacey (1967) specifically set out to examine the claim of the Berkeley *Productive Thinking Program* that it was training generalized problem-solving skills. To measure this objective, Ripple & Dacey used the Maier Two-String Problem. They correctly claimed that the training provided did not specifically relate to this problem. Because trained subjects reached solutions significantly earlier than non-trained subjects Ripple & Dacey claimed transfer of training effects. However, time taken to reach a decision, the criterion employed, was not one of high priority in Social Studies programs. Similar differentiation between training and evaluation tasks could be noted in the Treffinger & Ripple (1970) study where arithmetic puzzle tests were used. Noting the strong similarity between the format of the instructional materials and the criterion measures used in previous studies they utilized non-specific measures of transfer by using tests that were dissimilar in both content and format to the training materials.

*Note the similarity of ideas here with the three regions of reference covered by the Brunswick terms, central, proximal and distal, more recently elaborated by Snow (1974).

Overall, the few available results concerning transfer suggest that some problem-solving skills taught have been transferred to new situations (see Olton & Crutchfield, 1969; Robinson *et al.*, 1972). Generally, however, the criterion tasks used must be viewed as being different from but similar to those used in training. Robinson *et al.* (1972) classified two of their tests as 'remote transfer' and a third as 'immediate transfer' but even the remote tasks were less remote from the training provided than those used in studies by Ripple & Dacey and Treffinger & Ripple. No significant effects of the instructional program were found in the last named study on arithmetic problem-solving tests. While admitting these problems were rigorous tests of transfer, the authors argued their relevance as problem-solving tasks.

Experimental Design

Many investigators have indicated the need for further research to identify the specific factors associated with the changes in problem-solving behaviour*. Generally the experimental designs employed make it impossible to disentangle the influence of such factors as teacher, school, classroom environment and rate of presentation of material from the contribution of the particular training program. Taba (1966) found it impossible to control factors such as teacher training and the program actually used, so the study was regarded as an exploratory one. Others also had reservations about the strength of their findings (see Jenkinson & Lampard, 1957; Robinson *et al.*, 1972; Case & Fry, 1974; Covington, n.d.). With the Case & Fry study only one teacher was employed to teach the enquiry program to a group of students withdrawn from English classes. The fifth grade treatment group for the Ojemann & Campbell (1974) investigation was located in one school and the control in another. For the sixth grade level, both treatment and control were within the one school but taught by different teachers. A similar situation pertained to Butts & Jones (1966). Two classes in the school were involved in a program designed to enhance their problem-solving behaviours. Another two classes in the same school served as the control. With the Cousins (1963) investigation a control group was not used. In Elsmere (1963), the experimenter was also the teacher of the treatment group.

*It should be noted that not all researchers advocate the use of experimental designs. Snow (1974), for example, was not completely sympathetic to the notion of controlling significant variables arguing that such an approach produces artificial situations and unnatural behaviour. He also discussed problems encountered in generalizing from observations of a sample to a target population, and the difficulty of identifying treatment dimensions. He argued strongly for the adoption of quasi-representative and for naturalistic, quantitative case studies in educational research.

Another teacher taught the control. Only Wardrop *et al.* (1969) payed attention to measuring classroom environment, a factor that proved to be important in implementing a successful problem-solving program.

Charter & Jones (1973) and Shaver & Larkins (1973) pointed out the need to describe and to monitor what actually takes place in the treatment groups. Were the teachers actually doing what the experimenter thought they were doing? Unless this is established, treatments which are different on paper may be dismissed as being similar in terms of contribution to student performance, when, in fact, any planned differences may have been lost in the implementation phase. In practice there may have been no difference between the independent variables. Mason (1963), for example, expected one group of teachers to use materials as printed but another group was free to make modifications. The extent to which the first group conformed was not documented, nor was the degree to which the second group deviated from the planned teaching strategy. In the Ojemann & Campbell (1974) study, the teachers were encouraged to use their own experience and imagination in working towards the accomplishment of particular objectives. The approach of Jenkinson & Lampard (1957) was similar. Neither study detailed the program actually taught. Teachers in the Cousins (1963), Elsmere (1963) and Massialas (1963) investigations kept daily diaries as a record of what took place in class. Tape recordings were made of some class discussions to validate the daily log. In the case of Massialas, for example, this taping was done on two occasions. However, neither with Cousins nor Elsmere was the stated purpose of the taping to check the extent to which the teacher followed a planned procedure. Nor was the taping done to provide a description of the class activities. In both cases the taping was used to document the frequency with which students used elements of the enquiry process. With Taba (1966) and Massialas & Zevin (1967) interest was upon the interactions between teacher and students. Neither Taba nor Massialas & Zevin required teachers to follow a set program.

In terms of statistical techniques used to analyse data, two points need to be made about previous research. First, rarely were the assumptions underlying the use of particular statistical procedures acknowledged. Did the data justify the use of particular statistical procedures? The study by Wallen *et al.* (1969) was one exception to this criticism. There it was discovered that the gathered data did not match, in all instances, the underlying assumptions of analysis of covariance: normality, homogeneity of variance and homogeneity of regression. Consequently, the interpretations of the statistical analyses were considered as suggestive only. Secondly, few investigations have employed one or more of the variety of multivariate techniques

developed over the past decade. This oversight is unfortunate as multivariate techniques are eminently suited to classroom research, where complex processes are being examined and where many predictor variables and criterion measures are being employed (Finn, 1974; Keeves, 1974).

Implications for Further Research

From this review of research a number of points emerge which should be considered in planning further studies. First, training programs should be carefully articulated in relation to an underlying rationale. Results derived from certain activities can then be interpreted in terms of this reasoned explanation. Also, a rationale helps to identify appropriate measuring instruments; instruments to describe the treatment sample and to act as covariates in any planned analyses. Secondly, provision should be made to establish the extent to which the program as originally conceived was actually implemented. This would entail monitoring a sample of lessons to determine whether the procedures detailed for particular teaching episodes matched the teacher's operation. Thirdly, a variety of instruments should be used to assess performance on both the total problem-solving task and the separate enquiry skills. Solving the problem tasks should require the use of a variety of enquiry skills. These problems should not be segmented into a series of sub-questions by the tester as this changes the nature of the activity; the M-space requirements are changed. Assessment of performance should utilize multiple criteria. Problem tasks should be regarded as measures of transfer where they differ in content and format from the problems used in class. Fourthly, to determine the nature of the training effects, measures of both transfer and durability should be used. If genuine cognitive growth has been achieved it would be expected to transfer to new situations, and not dissipate after a few weeks. With the use of both transfer and durability measures a distinction could be made between improved performance and improved competence (Flavell & Wohlwill, 1969). Fifthly, an experimental design should be adopted which allows the factors influencing treatment success to be disentangled. Multivariate statistical procedures should be used to identify the extent to which different factors influence performance. Where such procedures are used the assumptions underlying their use should be examined.

As has been seen in this chapter, some of these features have been included in previous research. But no one study included all aspects. Most of the reported investigations were deficient in more than two attributes. The next two chapters detail how the present research considered and applied these guidelines. Chapter 5 describes the training program and how it relates to the Pascual-Leone view of factors influencing cognitive growth. Chapter 6 describes the experimental design, testing instruments and statistical procedures.

5. The Training Programs

The Problems

In Social Studies programs the problems investigated are usually characterized by numerous pieces of information and by the complexity of the relationships. There are large numbers of considerations relevant to any problem, and there is rarely an indisputable solution. Indeed, in the context of Social Studies, any problem not exhibiting these features might be considered trivial. Problem-clarification, the collection and consideration of information, and discussion of ways to present findings are activities normally spread over a period of days. Such a situation was not considered appropriate in the present training program. The duration of the program was strictly limited to six weeks. In that time it was not considered desirable for the student to become submerged in the specific details of one or two problems. Rather, it was hoped that students would develop an overview of the enquiry process and master some of the skills associated with gathering and manipulating data. It was assumed that these objectives would best be served if subjects faced many problems so that the relevant skills could be brought to their attention and practised on many occasions. To this end, the problems examined in the training sessions were not of a type that could necessarily be labelled as socially significant or of being the same as those normally examined in Social Studies programs.

Many writers in discussing problem-solving have emphasized that a question is not a problem unless a gap exists between what the person actually knows and the achievement of an appropriate solution (Dewey, 1933; Bartlett, 1958; Ausubel & Robinson, 1969; Evans, 1972). Dewey, for example, wrote:

> *Unless there is something doubtful the situation is read off at a glance, it is taken on sight; that is, there is merely perception,*

recognition and not judgment. . . But if it suggests, however vaguely, different meanings, rival possible interpretations there is some point at issue, some matter at stake.

(Dewey, 1933:121)

Not all questions meet this criterion. However, the questions used in the present investigation did give scope for the application of the abilities of analysis, synthesis and evaluation. Thus, the gap requirement of legitimate problems was met.

Generally, a characteristic of Social Studies questions used in research studies is their vagueness. The intent of the questioner is often unclear to the uninitiated. Palfrey (1972) examined some of the questions asked by Piaget of children to see how they perceived their world. Palfrey linked the vague, ambiguous nature of the questions to the cognate responses they elicited. Lunzer (1970) also commented upon the uncertainty associated with Piagetian-type problems. The questions used in the present study had a similar open-ended quality. However, unlike the Peel (1971) problems, they did not involve moral issues. Questions used included; 'From where do these people obtain most of their food?', 'How might these people earn money?', 'Why have these people chosen to build their houses on the water?', and 'Are there washing machines in this community?'

All the questions asked were based upon human activities portrayed in photographs. A wrestling festival in Mongolia, Navajo Indians living in a desert environment, and Kikuya women preparing cornmeal outside their hut were examples. For each problem, the data given suggested alternative solutions and there was scope for the subjects to relate it to similar situations in their own community. The subjects were not forced to go outside the information given to seek a solution, but neither were they restricted to the information given to achieve an answer. For each problem, a variety of plausible solutions existed. Data supporting each of these was available so there was scope for the subject to display his ability to resolve cognitive conflict.

With mathematical and physical science problems, investigated by many researchers, a precise identification of cognitive load requirements was possible because there was only one appropriate solution and the variables that had to be manipulated to gain it were known (Inhelder & Piaget, 1958; Pascual-Leone, 1970; Case, 1972*b*). As Furth pointed out, 'our knowing of and dealing with persons can never reach the level of abstract logical certainty that we can apply to physical reality' (Furth, 1970:129). Thus, in the case of the physical science problems, the appropriate executive, figura-

tive and operative schemes can be detailed more readily than is the case with Social Studies problems. At best, the adequacy of a response in a Social Studies setting can be assessed in terms of such factors as quantity of evidence used, whether alternative solutions were considered, and the care and systematic manner in which the final solution was identified. The classification of such responses is discussed in detail in Chapter 6. Yet it is clear that the cognitive demand of Social Studies tasks is beyond the M-space available to most primary and junior secondary students.

In the training sessions the problem to be solved was given to the subjects only after they had viewed the photograph to which it related. Primarily this procedure was adopted because of the work of Rothkopf (1966, 1970) and Frase (1970). Rothkopf maintained that questions aid in the acquisition and retention of information, but the particular outcome is influenced by where the question is placed in relation to the data with which it is associated. Pre-questions can act as filtering devices causing a subject to overlook information which in subsequent discussions may be useful. Rothkopf (1966) found that, in general, learners retained more of the given material when questions appeared after the data. Frase (1970) found retention of incidental information was relatively low when pre-questions were used. In some cases pre-questions depressed incidental learning well below control group scores (Frase, Patrick & Schumer, 1970). In the present investigation, by not nominating the topic of concern, the subjects were forced to scan and recall more completely the information presented. Thus, at this early stage in the enquiry operation, the object was to get the subject to identify as much information as possible and not to reject some because of a superficial examination. As such a procedure imposed a considerable information load upon the child, strategies were employed to aid him achieve the objective. These strategies are identified below.

Teaching Strategies

In addition to the questions which form the focus of the enquiry process, questions were also part of the teaching strategies used in the training sessions. Skill in question-asking has long been emphasized in teaching. More recently, however, the adoption of particular questioning procedures has been seen as important in improving intellectual performance (Pate & Bremer, 1967; Hunkins, 1970; Clegg, 1971; Taba *et al.*, 1971; Ryan, 1973; Victoria, Education Department, 1974). Taba identified four types of questions. They were described as opening questions, extension questions,

explanation questions and support questions. The opening or focus questions provide the boundary of subsequent discussion. They are identical to the problem questions discussed above. They require the child to elicit and manipulate a number of pieces of information. Extension questions are those which call for more information or clarification of what has already been provided. By contrast, explanation questions require more than description. Support questions are basically questions which ask the child to clarify and synthesize his ideas. All of these forms of questions were used in the training sessions associated with one treatment and to a lesser extent with the other special treatment. However, it is true to say that the teacher, irrespective of the treatment, had an important role to play in each training session. The subject was not left to work his way through a programmed text but rather encouraged by the teacher's use of questions.

Using Photographs

The training sessions used photographs as the source of data. These photographs, with few exceptions, showed human activity in cultural circumstances different from those of the subject's. Large, coloured, class discussion prints were selected from a number of commercially published sets. Two of these sets, *Contemporary Social Science Curriculum* and *Social Studies Discussion Pictures** were produced to accompany Social Studies programs designed for elementary school children in Canada and the United States of America. The third source was a picture series produced by UNESCO. Associated with each photograph was a problem task as discussed above.

The use of such data was apt considering the emphasis currently placed upon the use of visual materials in Social Studies both at the primary and secondary school levels in Victoria. In this regard, the study was different from the majority of those reviewed earlier which explored the impact of teaching upon problem-solving. In those instances verbal material predominated as the stimuli.

Visual materials are not viewed as accessories or 'spiritual crutches' supporting verbal forms of instruction. Justification for their use is varied. It

*Anderson, E. A. *et al. Contemporary Social Science Curriculum*, Park Ridge: Silver Burdett, 1972. Owen, E. *Social Studies Discussion Picture*. Vancouver: Fitzhenry & Whiteside, 1972. UNESCO Geography Series, Set 2 *Asia*. UNESCO 1970.

is commonly asserted that a picture is worth a thousand words. Although the relative value of these two means of communication is not detailed, it is accepted in some programs that it is easier for children to operate with information which is conveyed in a visual rather than a verbal form. (Renehan & Wilkes, 1973; Victoria, Education Department, 1974). These same courses emphasize the use of direct experience, that is, experience where the child is physically present in a situation, where he can actively manipulate objects. This position of valuing direct experience for primary school children, who in the main are concrete operators, was extrapolated from the work of Piaget (Flavell, 1963; Piaget, 1970). Because it is not always feasible or practical, in a school situation, to provide direct experiences, some programs advocate the use of photographs which depict aspects of reality. Such are seen as surrogates for direct experience. The more lifelike the photograph, and the greater its fidelity to the real situation, the more appropriate it becomes for primary school children.

Realism is emphasized not only because curriculum developers believe realism aids learning but also because Social Studies is concerned with investigating reality. An aim in Social Studies is to develop concepts and generalizations about the real world and not some emasculated or stereotyped situation. Through the development of ideas, such as those relating to social organization, natural environment, change and self, it is hoped the student will be better able to cope with, respond to, and act upon situations existing in his own society. As photographs can be a vehicle for portraying aspects of reality, they are seen as being important in modern Social Studies programs.

Another major justification for the use of photographs is their motivating influence. Photographs have been used as means of sparking interest in a topic or setting the scene for enquiry. Teaching units produced as part of *Society in View* (Victoria, Education Department, 1974) use photographs as confrontation experiences. These activities aim to arouse a child's curiosity, place him in a position where he needs to resolve cognitive conflict and focus his attention upon particular aspects of a social situation. Boast (1973) suggested photographs have other merits. They can be examined and re-examined at leisure and can be arranged readily in a sequence to reveal change.

The empirical evidence does little either to dismiss or to support the justifications advanced by Social Studies programs for using photographs (see Rhys, 1966; Murray, 1970; Whitehead, 1972). Their use must continue to rely upon propositions such as 'realism aids interpretation'. A few writers

have argued for realism, believing that it will make learning easier and more complete (Gibson, 1954; Miller, 1967). Their basic position is that the closer an illustration approaches reality the greater the number of cues it contains. More cues means the student has a better opportunity to differentiate the situation in a finer and thus more accurate manner. From such experiences the student would be able to build up more abstract and more inclusive concepts.

There is some evidence that challenges the theoretical position that realism in photographs aids interpretation (Dwyer, 1971; Moore & Sasse, 1971). Both Dwyer and Moore & Sasse found line drawings more effective in transmitting a message. Dwyer argued that photographs are less effective than line drawings in facilitating learning because they require more coding by the central nervous system. In addition, he suggested realism in photographs distracts students from the essential learning cues.

The justification for using photographs as a source of data in the present study rests on three grounds. First, photographs are accepted as an important form of information in Social Studies programs. Secondly, at least initially, they appear to attract the attention of students at the upper primary level. Thirdly, by 10 years of age most children have developed the elementary skills to comprehend the actions portrayed. (See Bayless & Renwick, 1966; Henderson, 1968; Vernon, 1968; Travers, 1969; Boast, 1974.) The evidence regarding the ease of photograph interpretation in relation to the comprehension of prose passages and direct experience is not clear. Nor has the importance of realism in learning been resolved.

Nature of Training

The training began after the initial random allocation of subjects to three groups and the completion of a pre-testing program. Two groups received special training, while the third acted as a control.

The training was spread over six weeks. During that time the training groups received 15 sessions. In one week they received two sessions while in the next week they had three. The sessions were taken on successive days with each session lasting one hour. Thus, a total of 15 hours experiences was provided for each experimental subject.

As experimental teachers worked in two schools, the division of training into two and three sessions per week allowed the teachers to spend whole days in each school. If, in week one, the teacher was working at School A on Monday and Tuesday, and School B on Wednesday, Thursday and Friday

then, the second week, he was at School A on Monday, Tuesday and Wednesday and at School B for the remainder of the week.

Treatment One

Treatment One was developed after considering the Pascual-Leone view of cognitive development. It aimed at providing suitable executive schemes and at enhancing appropriate operative and figurative schemes. Strategies were employed which aimed at reducing the cognitive demands of problems while at the same time encouraging each child to utilize all the M-space available to him and to avoid prominent but irrelevant perceptual cues. In summary, the training was designed to overcome some of the constraints hypothesized to inhibit pupil performance, while at the same time providing instruction in appropriate enquiry skills.

The first nine sessions concentrated upon particular operative schemes seen as being important in the total enquiry process. The totality of the problem operation was not tackled until the tenth training session. This approach differed from a number of the previous research studies where components were not treated separately (Olton & Crutchfield, 1969; Wardrop *et al.*, 1969; Gray, 1972). Justification for separating out relevant operative schemes relates to the limited cognitive processing space available to each subject. By separating the various elements of enquiry the amount of information to be considered was reduced. There were fewer procedures to be recalled and related to one another. Attention was directed to practising and mastering one outcome. Part of this mastery was a chunking procedure, a consolidation of information into smaller numbers of units. For example, in the generation of possible solutions to a given problem, the child was encouraged to consider and use three approaches: search the given information for clues; recall similar problems and their solutions from previous experience; and to guess. Ultimately these approaches become part of the one answer-generation scheme.

While the early sessions focused upon particular elements of the enquiry process, they built upon one another. For example, the photograph of boy wrestlers in Mongolia used as stimuli in a finding information session was used again in a session concerned with identifying evidence to support a particular problem situation. Thus evidence collected in one session was used to support suggested answers advanced in another. The same problems and stimulus material were re-used in different training sessions, although the focus of attention in each session differed. In one session the concern was collecting evidence, in another, suggesting answers to a problem, while in

another, manipulating that evidence to reach a decision. Thus, the total enquiry process was being built up gradually, logically, and interconnectedly.

For each session a detailed lesson plan was provided for the experimental teachers (see Whitehead, 1975). Each plan gave a brief statement about the purpose of the session. There was a summary statement of what had gone on in previous lessons. This revision was for the benefit of the children. A list of photographs to be used and of problem questions to be raised was also given. The procedure for the session was then presented in some detail. Often the actual words that could be used were given, although the teachers were permitted to vary this to gain spontaneity, provided the intent of the session was maintained*. Examples of children's responses collected from trial sessions were given. In these instances teachers were encouraged to substitute responses drawn from their own subjects. Also provided were examples of questions of the type described by Taba *et al.* (1971) as extension, explanation and support questions. At the conclusion of the session a review statement of the day's activities was given and some indication made of the focus of attention for the next session.

An examination of the training sessions in more detail indicates that the first two sessions were concerned with the manner in which the child oriented himself to the data. 'Here is a photograph. What does it tell you? How many things can you find in the photograph? List them down.' Each child wrote down what he could find. He then grouped the listed information into things that 'go together'. Finally he gave each group a label which reflected the common attributes. Certain regrouping may have been necessary. Such behaviour was emphasized by Taba who said it was $n activity of concept development. The same strategy was adopted in Australian programs (Queensland, Education Department, 1972; Victoria, Education Department, 1974). The distinction between observed and inferred information was also discussed.

Several advantages were derived from such a procedure. First, it enhanced the figurative schemes of the subject. From a photograph of a large modern city a child might group tall office buildings, shops, cars, trains and people under the label city. This drew his attention to the features of 'city' and hopefully, through discussion with others in the study group, enabled him to identify new attributes which he did not formerly consider. Such

*From an analysis of tapes made of three sessions with each teacher and each treatment group, it was clear that the experimental teachers followed the planned activities very closely.

features could relate to cities as centres of entertainment, government and culture. Secondly, the strategy alerted the child to the abundance of information presented. Such awareness was important in later aspects of the enquiry process, including the generation of alternative solutions to given problems. Thirdly, by grouping and labelling the gathered information, the subject was assisted in chunking the material into more manageable and retrievable units, thus making more efficient use of the limited available M-space. Or, as Bereiter (1969) expressed it, the action is a provisional information processing move which has the effect of preserving the pieces of information in a form so they can be assembled later. Fourthly, adoption of this strategy reduced the possibility of a student observing only a limited number of dominant features. The field-dependent subject was helped to adopt a strategy which would ensure a systematic search of the photograph. Systematic search and de-centering are not always evident in the behaviour of primary school children, as the research evidence indicates. For example, Suchman (1961) noted perceptual biases in the search patterns of grade five students which, he believed, seriously restricted success in later aspects of the enquiry process. Therefore, it was only after the children had searched, listed, grouped and labelled, that the problem to be investigated was given.

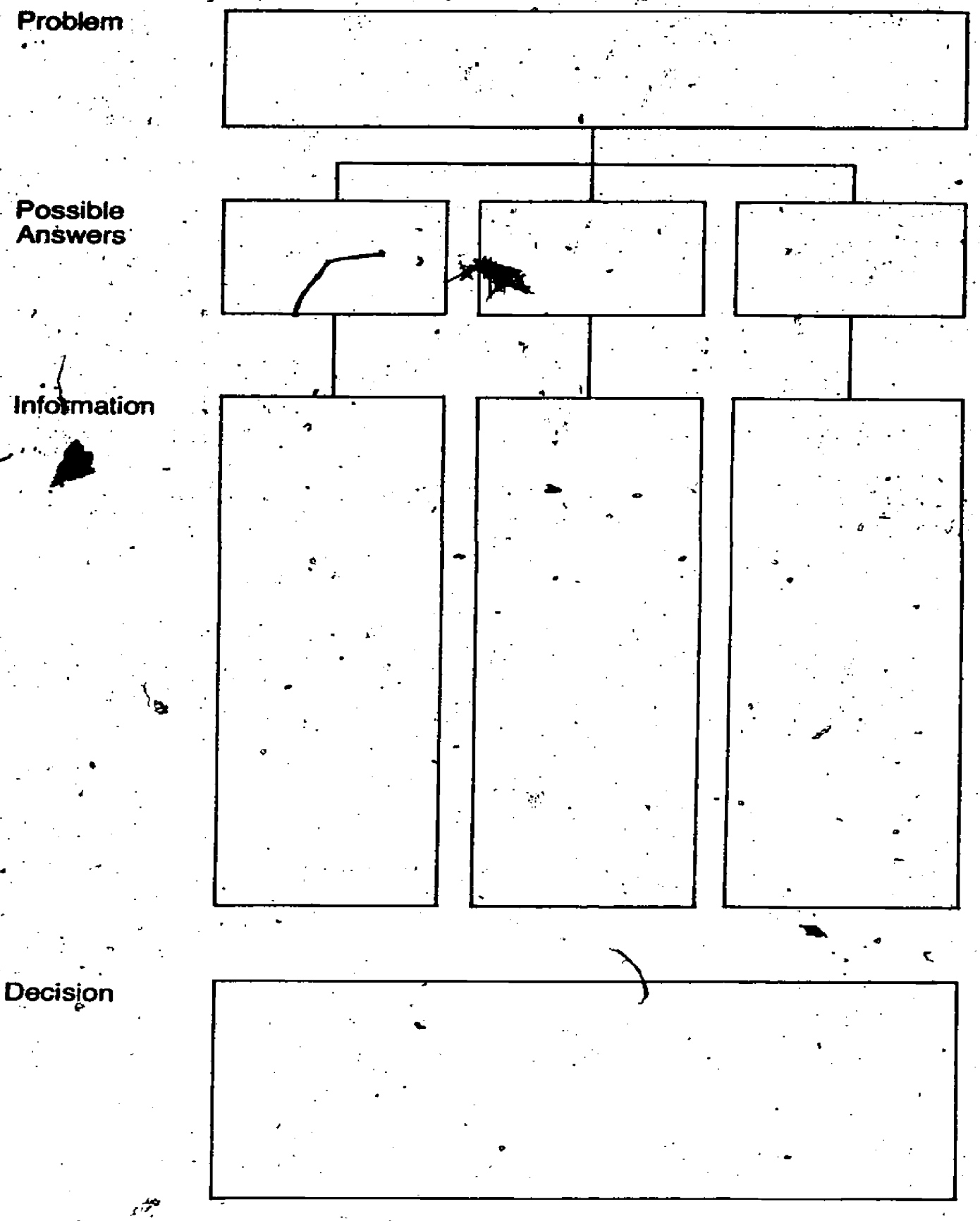
Sessions three and four were primarily concerned with finding multiple answers to given problems. This was the gap-filling phase of enquiry that many writers saw as being essential (Dewey, 1933; Ausubel & Robinson, 1969; Evans, 1972). Some have commented upon the difficulty students have in coping with this facet of enquiry (Lovell, 1961; Covington, 1968; Duckworth, 1972; Gray, 1972). These writers implied or suggested that the inability of students to provide multiple solutions to a problem may be due to the emphasis in classrooms upon single correct answers. In these two sessions the students were encouraged to view the problems in different ways. How can the problem be interpreted? No one suggested solution was valued more than any other, at least initially. Three possible sources of problem-solutions were given: something in the photograph, something seen or learned somewhere else, and something guessed. Use of imagination was encouraged. One piece of information may spark off an idea which appears to come from nowhere. In the group situation, it could be likened to brain-storming (Parnes & Meadow, 1959). Such ideas were as acceptable as those based upon previously held knowledge or upon given information. This feature or phase may be viewed as a creative act in the problem-solving process. So sessions three and four made explicit what suitable answers were, and gave directions to where such solutions could be found. Attention was therefore given to the development of suitable executive and operative schemes.

The next two sessions, sessions five and six, concerned themselves with data collection. This was the stage where the child returned to the photograph and listed information which supported or negated each of his suggested solutions. This was a more difficult task than simply listing information. The data collected had to relate to the given problem. 'What would be suitable evidence?' 'What does the problem mean?' It was the first step in explaining the appropriateness of a particular solution. Again, the emphasis was upon developing appropriate operative shemes.

Data manipulation was the focus of attention in sessions seven, eight and nine. Previous research had noted the limited nature of responses by students to problems of a similar type to those presented in the training sessions (see Chapter 2). Such limitations, as Pascual-Leone suggested, could be due to the inability of students to handle a sufficient amount of information or because they were unclear as to what was expected. These sessions presented a systematic way of approaching the task. The task was segmented into smaller units. The evidence supporting or negating each solution was considered in turn. One piece of information was set against another. The strength of the evidence between solutions was considered in a given sequence. Specific attention was given to nominating the attributes of successful decision-making: explanation rather than description, quantity of information used, and tentativeness in decision-making where evidence was insufficient or incomplete. Speed was not an important attribute. These procedures reduced the cognitive load. The task was segmented and information was chunked. Provision was made for the development of suitable operative and appropriate executive schemes.

Session 10 was the first of five sessions which concerned themselves with the total problem-solving task. Here was the first occasion where the various operative schemes used in earlier sessions were deliberately drawn together with the various relationships being identified. To do this, the crime situation, previously used in a more intensive manner by the *Productive Thinking Program* (Covington *et al.*, 1972), was used. By discussing the actions of the police on the scene of a crime the chief elements of enquiry were established: look around and establish the nature and setting of the crime or problem, clarify the nature of the crime, identify possible suspects or solutions, collect evidence or alibis for each suspect, make an arrest or reach a decision after weighing up the evidence. In addition to discussing the crime model, each child was given a duplicated paradigm upon which he recorded the various pieces of information gathered in each step of the enquiry. (See Figure 1.) This paradigm was as used by Robinson *et al.* (1972), although in

FIGURE 1 **Enquiry Paradigm**



that instance the subjects do not appear to have used it as a record sheet. This sheet served two purposes. It gave a visual image of the various tasks in problem-solving and their relationships, thereby assisting in establishing an appropriate figurative scheme of the total process. In addition, it provided a memory crutch and thus lightened the cognitive load. Once one section was filled in, it was temporarily set aside while another section was processed, thereby aiding the segmentation of the task.

While the paradigm form used identified the chief components of enquiry it also had built-in constraints. Provision was made for only three suggested solutions. Space limitations dictated that only three or four pieces of evidence could be included for each solution. Thus the amount of information listed came closer to the M-space available to the subjects included in the study. The subject was not choked with information half-way through the problem-solving process. The amount of information was controlled by the record sheet. While the child may have created many possible answers, he was told to select and list only three. When discussing their solutions the children were encouraged to consider the quality of the evidence. However, the basic emphasis was upon quantity rather than quality, that is, using all the evidence they had discovered. After the thirteenth session the paradigm was discarded and the children made rough notes on plain scrap paper.

Treatment Two

Primarily, Treatment Two differed from Treatment One in the extent to which directions were given to students. The teacher made no effort to clarify the nature of the problem-solving task nor to indicate the appropriateness of responses. The teacher acted as a discussion controller, encouraging each child to talk about the given problem. As a result, any improvement in student performance came from the child's own appreciation of the situation, and through observing the techniques of other students, not by matching the criteria emphasized by the teacher (as was the case with Treatment One).

It will be recalled from the discussion in Chapter 3 that Piaget believed students needed to be actively engaged in any intellectual task if they were to benefit from it (Piaget, 1970). Students should be given the opportunity to discover for themselves. Explanation or demonstration by teachers was not seen as being compatible with his views. Tasks should be provided which cause conflict in the child's mind. In resolving the conflict through the processes of accommodation and assimilation intellectual gains would be made. The interaction of ideas between peers would assist the process (Piaget, 1972).

No attempt was made by the teachers to give students in Treatment Two an overview of the enquiry process. The crime model was not discussed. The special recording sheet presenting a graphic view of the process was not used. Instruction in the separate skills associated with enquiry, such as generating answers and evaluating solutions, was not given. The teacher adopted a low profile in terms of the student group. Each subject was left to himself to identify and use the facets of enquiry. However, identical problems to those given to Treatment One were considered.

With one major exception, the training sessions experienced by Treatment Two subjects reflected features advocated by the Piagetian view. The exception related to the first phase of the enquiry process, data-scanning. Sessions one and two were identical for both Treatments One and Two. In those sessions the students were instructed in listing, grouping and labelling data. Perhaps this section, more than any other, emphasized the content matter of the enquiry—the clarification and enhancement of figurative schemes or concepts. In some modern Social Studies programs it is the development of this phase of enquiry that receives strong emphasis. Subjects in Treatment Two were therefore not placed at a disadvantage compared to Treatment One in terms of problems seen and scanning strategies employed in relation to photographs. However, teacher assistance went no further.

Treatment Three, The Control

The third group acted as the control. For that group the treatment consisted of listening to a story read by the experimental teacher. The story selected did not relate to the special treatments nor to the subject matter of the criterion tests. Both experimental groups had the same story read to them but for a much shorter duration.

Table 2 presents a summary of the attributes of the experimental treatments for the benefit of the reader.

The next chapter describes the experimental setting and design of the investigation in which the three treatments were employed.

TABLE 2
Similarities and Differences Between Treatments

Treatment One	Treatment Two	Treatment Three
1. Concept development *listing *grouping *labelling	1. Concept development *listing *grouping *labelling	
2. Explicit instruction *suggesting solutions to problems *finding evidence *manipulating data		
3. Discussing problems with teacher guidance and direction	2. Discussing problems. No teacher instruction	
4. Using a record sheet		
5. Listening to a story	3. Listening to a story	3. Listening to a story

6. Design, Sampling and Experimental Method

The present investigation aimed at examining the influence of training upon problem-solving performance. While the previous chapter described the training programs used, this chapter sets out the experimental design, the sample and kinds of data collected.

Design of the Study

The experimental design is shown in Table 3. Treatments were crossed with classes. Classes nested in schools and schools nested in teachers. By using such a design, the observations of treatment groups were independent of each other.

As Table 3 indicates, a total of 216 subjects formed the sample. Twelve classrooms from six schools were involved. From within each classroom, 18 subjects were randomly selected to take part in the investigation. Once identified, these students were randomly assigned to one of the three experimental groups: Treatment One, Treatment Two or Treatment Three. Thus, within any one school there were 36 subjects involved, drawn from the grade six level and equally divided between three experimental groups, each experimental group having 12 subjects.

As Table 3 shows, each experimental teacher was involved in teaching at two schools. Specially employed from outside the school staff, these experimental teachers taught all three experimental groups within each school. So each treatment group had the same teacher and each group was made up of students drawn equally from the same two classrooms. Different teachers for different treatments were not used nor were the students from different classrooms instructed separately.

TABLE 3
Experimental Design

			Treatments 1	2	3		
Teacher 1	School 1	Class 1	6	6	6	18	36
		Class 2	6	6	6	18	
	School 2	Class 3	6	6	6	18	36
		Class 4	6	6	6	18	
Teacher 3	School 3	Class 5	6	6	6	18	36
		Class 6	6	6	6	18	
	School 4	Class 7	6	6	6	18	36
		Class 8	6	6	6	18	
Teacher 3	School 5	Class 9	6	6	6	18	36
		Class 10	6	6	6	18	
	School 6	Class 11	6	6	6	18	36
		Class 12	6	6	6	18	
			72	72	72		216

Half the total sample was pre-tested on a set of criterion measures. This sub-sample was gained by randomly selecting half the subjects from each classroom within each treatment group. This can be related to Table 3. From each treatment cell of six subjects, three were pre-tested. This procedure was adopted to provide a check upon the composition of the treatment groups. Were they initially equivalent? As the three treatment groups were to be compared on tests given at the end of the experiment, the groups would have to be equivalent at the beginning of the experiment if comparisons were to be meaningful. However, the experience of doing the tests could have constituted a learning situation in itself. Perhaps the training sessions were contributing nothing more than was gained from the experience of doing the tests. Test-taking could be alerting the appropriate executive schemes and that may be the sole contribution of the training sessions. By pre-testing only half the sample a comparison was possible between those pre-tested and those not. Such a comparison would identify the extent to which the test-taking experience was a learning experience. The relationship between pre-testing and post-testing is illustrated in Table 4.

TABLE 4
Testing Sessions

	Pre-test		Post-test	Delayed Post-test
Treatment One	Tested	Training	Tested Tested	Tested
Treatment Two	Tested		Tested Tested	Tested
Treatment Three	Tested		Tested Tested	Tested

Also from Table 4 it can be seen that the half of the sample not pre-tested were re-tested after the post-testing session. Eight weeks was the planned time lapse between the completion of the post-testing and the commencement of the delayed testing session. Two of those weeks were school holidays. Use of a delayed post-testing session gave an opportunity to examine the durability of any identified training effect.

This discussion on pre-, post- and delayed post-testing relates to a group of tests referred to as Problem Tasks. Six in number, they are described in detail below. Three used photographs and three were based upon verbal passages. Four only, however, were used at any one testing session; two photographs and two verbal passages. The four tests used in the pre-testing phase were identical to those used in the delayed post-testing session. Two of these tests were used in the post-testing session along with two others. Thus each subject, irrespective of whether he was in the pre-test or delayed test group, completed six problem-solving tasks, two of which he saw twice. This relationship between tests and testing sessions is shown in Table 5.

TABLE 5
Testing Sessions and Problem Tasks Used

Pre-testing	Post-testing	Delayed Post-testing
Photograph One	Photograph One	Photograph One
Photograph Two	Photograph Three	Photograph Two
Passage One	Passage One	Passage One
Passage Two	Passage Three	Passage Two

In addition to the Problem Tasks, some tests were given to all subjects prior to the training sessions. These were tests which helped to describe the sample and treatment groups, while at the same time providing measures, independent of the special treatments, which could be used as covariates in the data analysis. These tests were *ACER Intermediate D*, a test of general ability, *ACER Word Knowledge Test Form B*, *Weschler Backward Digits Test*, and the *Group Embedded Figures Test* by Oltman, Raskin & Witkin (1971). Jones (1972) showed that performance on Piagetian tasks did not relate to verbal ability; however, in the criterion tasks employed in the present investigation such ability could have been a factor significantly influencing performance. Three of the Problem Tasks used data consisting of words. Furthermore, the child's response to these problems was given verbally in an interview situation rather than through the manipulation of materials. Johnston (1972) and Whitehead (1972) did find a significant but low correlation between these types of tasks and measures of intelligence. The *Backward Digits Test* and the *Group Embedded Figures Test* were included because they were measures which related directly to the Pascual-Leone view on cognitive development.

Besides the Problem Tasks, the post-testing session included measures described as Element Tests. They were so labelled because they aimed to measure particular attributes of the training program. Focusing on more limited aspects of the enquiry process, it was anticipated that they would assist in identifying changes in the underlying operative schemes that help to explain any change in performance on the more global problem-solving tasks. Two published tests were also included in this phase of the testing program. They were two tests specially prepared by the Australian Council for Educational Research to measure changes in the performance of primary school children in grades four through six as the result of introducing a new Social Studies course of study in Victoria. These tests were entitled *Test of Words used in Social Studies* and *Test of Comprehension*. All the tests used are listed in Table 6.

In summary, the investigation involved a sample of 216 students drawn from 12 grade six classrooms within six schools. Each of three experimental teachers, teaching all three experimental groups, worked in two schools. The nested design allowed the separation of the variables—school, teacher, and classroom—from treatment effects, factors not always considered in previous research. The adoption of a testing program which included pre- and delayed post- testing with a split sample permitted the examination of durability of training and the influence of test-taking upon performance.

TABLE 6
Tests Used in Investigation

Pre-testing	Post-testing	Delayed Post-testing
5. Four Problem Tasks*	1. Four Problem Tasks	1. Problem Tasks*
a. *Poverty*	a. *Poverty*	a. *Poverty*
b. *Slums*	b. *City*	b. *Slums*
c. *Indian Horses*	c. *Indian Horses*	c. *Indian Horses*
d. *Goldfield*	d. *Famine*	d. *Goldfield*
1. *Intermediate D*	2. Element Tests	
	a. Concepts: *Poverty* and *City*	
2. *Word Knowledge B*	b. *Answers*	
3. *Backward Digits*	c. *Puzzles*	
4. *Group Embedded Figures Test*	d. *Working it Out*	
	e. *Higgins Inference Tests*	
	3. ACER Social Studies	
	a. *Words used in Social Studies*	
	b. *Test of Comprehension*	

*Administered to half the total sample.
Numerals indicate the sequence of administration.

The Sample

Two hundred and sixteen grade six students were selected from six primary schools. The schools were drawn from those nominated by Education Department administrators and from those known to the investigator. All schools selected met four criteria.

They were located in middle-class areas.

There were no serious language problems in the school. All children included had a ready command of the English language.

Space was available in the school where a group of 12 children, together with a teacher, could go for the purpose of instruction and discussion.

The school principal and class teachers were agreeable to taking part in the investigation. This entailed providing time for the extensive testing program and allowing groups of children to be withdrawn throughout the school day, for instruction by the experimental teacher.

In the opinion of the class teachers, no child in the study had difficulty in expressing himself or herself in the English language, although 31 subjects came from homes where another language was spoken by at least one parent. Support for the teachers' opinion was reflected in student performance on a vocabulary measure and a test of general ability which had a high verbal component (see Whitehead, 1975).

The subjects were drawn equally from two classrooms in each school. With the exception of one school these were the only classes at grade six level. The exception was School Five which had three classes. In that instance two were selected at random. Generally, the environment of all classes in the investigation reflected traditional organization, one teacher being responsible for each class. Specialist staff were provided for work in an art and craft room and in the library in all but two schools. Both lacked special art and craft facilities but had libraries.

Since the major research hypotheses related to the different treatments and their possible effects, and since a full pre- and post-test design was not employed, it was essential that the membership of the treatment groups be made effectively random. School, class and teacher differences could arise since membership within and across these groups was not necessarily random.

Four tests were administered prior to the application of any treatments. *Backward Digits* and *Group Embedded Figures Test* were selected because they related to two constraints upon intellectual performance nominated by Pascual-Leone. The remaining two tests were *ACER Intermediate D*, a general ability test, and *ACER Word Knowledge B*. The general ability test was used because it was likely to have a positive correlation with the various criterion measures concerned with treatment effects. The *Word Knowledge Test* was included because three of the Problem Tasks used verbal passages.

These four tests permitted verification of the practical effectiveness of the procedures used for design and random selection of the samples. A series of simple univariate analyses of variance were conducted using the scores on these four variables. The F values for the treatment groups on all four variables were not significant ($p<.05$). For class, school and teacher effects

some F values were significant. These results were consistent with the effectively random selection of subjects across treatment groups.

In addition to the administration of the four tests, half of the sample were pre-tested on four Problem Tasks. These were the major cri[illegible]asures for the investigation. An examination of the pre-test resu[illegible]ed the conclusion drawn from performance on the other instrume[illegible]ere was no significant difference between treatment groups.

Tests

Five major groups of measures were used in the present study. They are listed in Table 6. One set, Problem Tasks, examined the ability of students to use the total enquiry process. Since the second set related to separate components of the process they were grouped under the general label, Element Tests. These tests were measures of 'take' (Robinson *et al.*, 1972). The third group consisted of measures specifically associated with the Pascual-Leone framework of intellectual development. Published objective tests made up the fourth group. The final group of measures provided checks on classroom environment and the extent to which the planned teaching procedures were actually implemented. The measures involved in each of these five groups are discussed below.

Problem Tasks

Description

There were six Problem Tasks (see Whitehead, 1975). These were titled *Indian Horses*, *Famine*, *Goldfields*, *Poverty*, *City* and *Slums*. The first three problems were based upon written passages. The last three used photographs; one black and white, and two coloured.

The verbal passages were built upon the model proposed by Peel (1966) in which the first section was basically irrelevant to the problem posed. The next section, although related to the question asked, was in itself inadequate or insufficient for a conclusive judgement to be made. The final section provided or implied other circumstances that limited the nature of the decision that could be reached. Each of the passages could be viewed as either history, geography, ec[illegible] sociology or anthropology.

The first problem concerned the original source of the horses used by the American Indians. 'From where did the American Indians obtain their horses?' This was an issue that at one time interested American anthropologists (Haines, 1938). The second problem focused upon famine in India and the dilemma faced by those people in matching religious beliefs

with the economics of survival: 'How can the shortage of food in India be overcome?' The third situation was prompted by the Lambing Flat Riot of 1861 between European goldminers and Chinese. Highlighting differences in cultural and social backgrounds, the problem focused on procedures that could have been used to diffuse the animosity between the groups. The question was 'How could trouble on the goldfields be prevented?'

The photographs were of the type found in Social Studies texts and classroom photographic sets. The first photograph portrayed a Mexican village. It showed crude houses built of brick but lacking in facilities. A woman was washing outside in tubs. Neglect was apparent in the surrounding area. With this photograph, the subjects in the study were told that the government of these people in recognizing they were living in poverty had provided money for a community project. 'What should the project be?' was the problem posed. The second photograph was an engraving of the building of the overland telegraph line between Adelaide and Port Darwin in the late 1800's. Tents were erected alongside a creek and considerable activity was shown. Men were leading pack horses, erecting telegraph poles, carrying buckets of water and cooking. Although a rock outcrop was featured in the foreground, the land was reasonably flat and covered in light scrub. Transport was available. The question asked was 'Will there be a city here, in ten years time from when the drawing was made?' The third photograph showed a section of Buenos Aires. Separated from high-rise buildings in the background by a railway line and railway stock, the foreground showed small wooden houses with corrugated iron roofs. While mainly old and dilapidated, some houses showed signs of care. With this photograph the students were told the local Council had decided to pull down these houses and asked 'Why has the Council decided these houses must be pulled down?'

All six situations provided a large amount of information. The Pascual-Leone M-space model would suggest the amount of information provided was beyond the processing space available to grade six students. Yet the problems were of such a nature that a response could be made by all students irrespective of their level of intellectual maturity. Unlike many Piagetian problems, there was no pre-determined correct solution that could only be gained by manipulating certain variables. Rather, there was a variety of acceptable answers, whose quality was only conditioned by the amount of evidence used to support them, and the care with which the appropriateness of the information was sorted and acted upon. A more detailed discussion of response rating is given below. Each problem situation gave scope for the application of almost all aspects of the enquiry process. Students could

clarify the nature of the problem, seek possible solutions from both within and outside the given data, gather evidence to support each alternative answer and evaluate this evidence in terms of each alternative, and so make a decision. The one ability that was missing was problem identification. In all cases, the problem to be investigated was pre-determined by the experimenter. However, there was scope for problem clarification.

Using the Robinson *et al.* (1972) dimension of remoteness, the visual tasks corresponded closely to those used in the training sessions. Two of the situations actually used the figurative schemes that were featured in the training sessions, namely, poverty and city. So these measures were towards the 'measure of take' category identified by Robinson *et al.* The verbal tasks, although using the same skills and using Social Studies-type situations, were more remote from the child's experience. No attempt was made in the training sessions to instruct the child in applying enquiry techniques to written material. Therefore, the verbal tasks could be viewed as measures of transfer. However, since the problems were still within the Social Studies range, rather than say the physical sciences, the tasks must be considered as 'immediate transfer' rather than 'remote transfer' as was the case with the Maier String Problem used in the study by Treffinger & Ripple (1970).

All the Problem Tasks were administered in an interview situation. The attention of the student was drawn to a pad and pencil provided. He was free to use them if he wished. It will be recalled that the training sessions emphasized the use of brief notes as a memory aid. Each problem was presented after the given data was examined. In the case of the verbal passages, these were read aloud by the tester as the subject followed with his eye. The data was left in front of the subject as the relevant question was posed.

Trial of Tests

All six Problem Tasks used in the present study were tried out with 48 subjects, selected at random, from three grade six classrooms. This trial testing of the Problem Tasks indicated four things. First, all children were able and willing to respond to the situations presented. Secondly, the level of response by most students was well below the ceiling of the test. Only two children out of 48 reached the seventh category of performance on a scale providing for 10 categories. Therefore, the instruments provided gave scope for recording any improvement in performance arising from special learning experiences. Thirdly, although the tasks drew from different content areas and utilized two different modes of data presentation, there was a reasonable level of shared variation between the tasks, sufficient for them to be viewed

as components in a common set of measures. Fourthly, the instruments were reasonably stable in the test-retest situation. Any significant improvement in performance of subjects after training could be attributed to the treatment effects rather than to inadequacies in the measuring instruments or benefits derived from previous experience with the tasks.

Marker Reliability

In responding to the Problem Tasks described above, highly complex answers were created. Such answers present difficulties for markers attempting to achieve a high level of reliability. In previous research, at least two approaches to the problem were used. One searched for particular elements in each response, acknowledging individually their presence with separate scores. These scores were then summed to give an overall measure (Shulman 1965; Olton & Crutchfield, 1969; Nelson & Abraham, 1973). The second approach was more impressionistic. Here the rater formed a general impression of the response in terms of a multiplicity of criteria. Usually these judgments were made in terms of a number of broad categories each having its own criteria. Generally this global approach was seen to be less reliable that the segmented approach. However, some researchers have believed that the global method reflects more accurately the essential nature of problem-solving. For them, problem-solving was not a string of separate skills (Peel, 1971). Rather it was an integration of skills to form a new entity. Thus the totality of the process was seen to be greater than the sum of its parts. While giving the appearance of precision, the summing of separate element scores could result in a distorted view of a person's ability to use the enquiry process. Obviously, the child who identifies 10 features in a problem and proceeds to operate upon two is performing quite differently from a child who identifies six features and operates upon all six. A partial solution would be to give more value to particular attributes of the performance to reflect the importance of each in the total process (Olton & Crutchfield, 1969; Robinson, n.d.). But in doing this the subjectivity of the global method is approached. Another solution, as adopted by Tunstall (1969), is not to combine the separate scores. However, by doing this the wholeness of the problem-solving process is overlooked, the ability to combine various skills to form new patterns of gap-filling behaviour being ignored.

Is it possible, using global methods of assessment, to achieve high levels of marker reliability? Peel (1971) and Nettle (1975) have provided evidence to support the claim that high levels of marker reliability are possible. Peel reported correlations in the order of 0.89 to 0.98 from a number of studies carried out by his students. In the main, three broad categories of behaviour were used. Each had multiple criteria. While a larger number of categories

were indentified in Peel's earlier work they were telescoped into them in order to achieve 'reproducibility as required by Guttman's technique for discovering whether any chosen set of grades constitute a scale' (Peel, 1971:39). However, Peel was still prepared to use sub-categories should the responses from particular test situations demand it. For example, the responses on the Pilot question were rated on four categories while those on 'Jane' have five categories. Obviously, the fewer the categories, the less precise the scale becomes in measuring the various attributes of the enquiry process. This is particularly important when a researcher is concerned with measuring change in performance due to the introduction of particular learning experiences. The dual requirement of precision and sensitivity in a scale present a dilemma. Improvement in one tends to worsen the other, so a compromise seems inevitable.

The six Problem Tasks used in the present study were tested with 48 grade six children. Their responses were rated against a nine-point scale described below. Using this scale, two raters achieved a marker reliability correlation of 0.94. In the study proper, one third of the transcripts were randomly selected to be assessed by two raters. On that occasion a correlation of 0.89 was achieved. The high correlations could be attributed to the systematic and common procedure adopted by the markers. Has the subject considered multiple answers? How much information has he given to support these answers? Did he go outside the information presented? To reach a decision, did the subject systematically weigh up the evidence for each alternative answer? Was there recognition of inadequacies in the data provided? Consideration of such questions helped to place the responses into appropriate categories in the scale. Overall, the procedure allowed for a high level of precision in judging responses.

Scoring Responses

Part of the marker reliability issue related to whether levels of ability in problem-solving could be clearly identified. In Chapter 2, the nature of the enquiry process has been discussed. There, different views and points of emphasis about this intellectual operation have been noted. However, in the context of the present investigation the enquiry process was seen to concern:

> the clarification of a problem or conflict situation;
> the creation of solutions to resolve the conflict;
> the gathering of data to support each possible solution, and
> the careful and systematic sorting and weighing of the data to reach a decision.

Evidence regarding the existence of these attributes in student responses was necessary for the classification of 'successful Social Studies problem-solver'.

Four broad categories of problem-solving behaviour were identified. Three of these groups were sub-divided to provide a nine-point rating scale. The first category related to the non-problem-solver. He did not see the nature of the problem. He made no response to the conflict presented or he gave an answer which bore no relationship to the given problem or the information provided. Occasionally he threw information together in a haphazard manner to produce a nonsensical statement. Another variation was where he created his own data to justify a particular response. Sometimes an answer was a tautological one: 'he is hungry because he is hungry'.

The type of behaviour placed in the first category should not be confused with the tentativeness seen in the higher level responses. Hesitancy exhibited in such cases was because the subject saw limitations in the data provided or because he recognized the breadth or complexity of the problem. Usually he gave an answer when asked but phrased his reasons in a manner admitting probability: 'It is most likely that. . .', or 'On the limited information given here. . .' or 'If this is true then it is possible that. . .'.

The second level of response encompassed the first act of the problem-solving procedure. Here the problem was seen but little mental effort was expended in solving it. The problem was comprehended sufficiently for the subject to give a sensible answer. However, attempts to relate the problem to given data were minimal. Two sub-levels could be noted. One covered the responses where a subject failed to recognize the existence of multiple solutions to a given problem. Here the subject seized upon *one* piece of information provided in the data to support a given answer. The second sub-category concerned those responses where the subject saw multiple answers to the problem posed but did nothing about settling on one. On most occasions he made no attempt to relate the given information to any of his suggested answers. This form of response was seen as being slightly better than one in the previous sub-category because it assumed the subject had not only recognized the existence of a problem but also had some awareness of its complexity. All the alternatives the child produced need not necessarily have come only from within the given data. Some may have arisen from personal experiences because the problems were not unfamiliar to the majority of primary school children. Famine, slums, Indians, gold-mining and cities were all ideas that have been raised in private reading and television viewing, if not in classroom activities. Perhaps a slightly better

response than the 'alternative answers only' was the one where two solutions were considered along with one piece of information. However, it too was rated in the second sub-category of the second level of response. Justification for doing this rested on the belief that both situations constitute a similar cognitive load. Three alternative answers, or two alternative answers plus one piece of information were equivalent, in that they demanded the same amount of mental processing space.

The main characteristic of the third response level was the collection of data to support possible solutions to the Problem Tasks. These responses were usually descriptive in nature. They relied, in the main, upon information provided in the test passage or photograph. Rarely did subjects classified within this level cite evidence from other sources. Generally no attempt was made to explain the conflict within the information presented to them. Where it was done, it was only done in a groping or experimental sort of way. Therefore such efforts were usually unsuccessful.

Within the third category four sub-stages were identified. Basically the differences between these response classifications were reflected in the amount of information used, the number of alternative answers considered and a tendency to pay attention to conflicts in the data. The first sub-level covered those responses where *two* pieces of information were used to support *one* answer. Here the subject expressed two ideas gathered from the given information to help him 'tell' his answer.

The second sub-level differed in that alternative answers were considered, although the information given in support of them was restricted. Along with *two* answers, *two* pieces of information could have been provided. Alternatively a student could have used three pieces of information to support one answer, the second answer, although identified, having received no support. These instances had more 'thought units' than those classified at the second sub-category of level two. There two alternative answers, together with one piece of information, was sufficient.

In the third sub-level of category three, only one answer was considered but *more* than two pieces of information were given in its support. Again no attempt was made to weigh this information up. It was presented in an unintegrated fashion.

Like sub-category two, sub-category four saw the suggestion of alternative solutions to a problem, although here *three* or more alternatives were presented. Supporting each possible solution was a piece of data. Responses which covered two possibilities, with several pieces of information for each possibility, were also placed in this sub-category. The justification for doing

this related to cognitive load. It was assumed that a response made up of three alternative answers and three pieces of information was somewhat equivalent in terms of mental effort to one where two answers were advanced along with four pieces of information. In all of the cases so far mentioned in sub-category four, there was no attempt to tackle the conflict evident in the data when related to the Problem Task. There were responses, however, which did try, in a tentative way, to do this. They, too, were rated at this sub-level. In such cases there was an attempt to relate pieces of apparently conflicting information together. What could have been seen as a disadvantage was neutralised or accommodated in the offered solution. An example of this type of response was seen in the *Famine* question. There the student was told that the Indians, despite shortage of food, were not prepared to kill cows wandering through the village. This was because of their religious beliefs. For a response rated at sub-category four the dilemma was resolved by getting 'somebody else to kill the cows', then the Indians had food and apparently their religious principles were not offended. This type of response was considered to be superior to those in earlier sub-levels because it gave evidence of data manipulation. While the response may have contained less information, the processing demands were greater. The response was labelled as 'tentative', 'unsystematic' and 'relatively unsuccessful' because only *one* compensation was evident. More than one compensation was considered worthy of a higher classification.

The fourth and final response level included the characteristics of the 'successful' Social Studies problem-solver. A major attribute of the responses classified at this level was that they demonstrated the subject's ability to analyse and evaluate data. With any problem, the subject operating at this level usually made a conscious effort to establish its nature and the criteria by which a solution was to be judged. In seeking solutions he was not bound by reality but included hypothetical possibilities. He looked beyond the surface features of a situation or the dominant elements. Once identified, the possible solutions were carefully and systematically related to the given data to eliminate the unprofitable. He reasoned by implication and attempted to reconcile conflicting information. In the process of doing this the subject recognized the inadequacies and inconsistencies within the data provided. Because of this he tended to couch responses in tentative terms.

These attributes were evident to varying degrees in different responses. Obviously various combinations were possible. In general, however, three sub-categories were identified which reflected increasing levels of complexity in the processing of information. The first included those responses

where a number of solutions were advanced and where attention was given to 'weighing up' the gathered data. The solutions themselves need not have arisen from the passage or photograph which formed the source of data. They could have been generated from the subject's personal experiences from the past, or from an analysis of the nature of the problem. Yet irrespective of where they came from, they were, each in turn, related to the information available in the source material. In processing the data, one piece of information was set against another to justify the acceptance of a particular decision or the rejection of another. To be classified in this sub-category there had to be at least two of these compensations; one was insufficient. However, all the alternative solutions may not have received a thorough examination. One may have got more careful scrutiny than the others.

It was in the second sub-category that all alternatives were carefully examined in terms of the data gathered. Responses in this category indicated that the subject saw the problem in a broader context than subjects classified in the previous sub-stage. Here there was much more than just the evaluation of evidence. In making his response, a subject may have expressed doubts as to the accuracy or completeness of the data provided, and because of this, he modified the nature of his final answer. In making this answer he often started speaking in general terms and then moved to specific pieces of information to support his argument. Similar characteristics were noted in the final sub-category. However, the final sub-category was distinguished from the previous level by the completeness of the response. Here the meaning of the problem was clarified, terms were defined and appropriate criteria for judging the adequacy of the response established. Almost all the evidence presented was manipulated in a systematic way, the apparent conflicts noted and attempts made to resolve them. The final solution usually indicated where additional evidence was required.

A summary of the criteria used to rate responses is presented in Table 7.

Element Tests

Five Element Tests were used in this study. Four were specially developed for the project. One was a test designed by Higgins (1974). They had two purposes. The first was to identify the level of student proficiency with certain skills seen as being important in problem-solving. In this regard these tests assisted the interpretation of results gained on the Problem Tasks described above. The second was to provide a means of measuring the success of specific attributes of the training program.

TABLE 7

Some Criteria Used to Assess Student Responses To Problem Tasks

Level	Attributes of Sub-categories	Score
I Jumbled and contradictory No sense	No response or 'reason' not from given data. Answer bore no relationship to the problem or the data—an imaginative response—fantasy. Lacked comprehension of the data. Tautological answer. Information combined in a nonsensical fashion—full of contradictions.	0
II Problem seen but little mental effort used to resolve it. Nature of problem comprehended. Response 'made sense'.	A. Did not recognize the existence of multiple answers. Seized upon one piece of information to support answer. Information derived from given data.	1
	B. Multiple answers advanced but little attempt made to relate given answer to the information provided. Two answers plus one supporting explanation. Three answers plus one piece of supporting information.	2
III Answers were basically descriptive. No attempt made to explain conflicting information or where an attempt was made it was unsuccessful. Given data provided sole source of information.	A. Two pieces of information to support answer—expressed in student's own words—involved translation.	3
	B. Alternative answers considered but information given in support of them restricted. May have included negative element. Two solutions plus two reasons. Two solutions with three reasons.	4

Level	Attributes of Sub-categories	Score
	C. More than two pieces of information relevant to one answer. No attempt to weigh up the information. Alternative answers not considered.	5
	D. Three answers with at least one piece of information to support each suggestion.	6
	Two answers each with more than one piece of data. Conflicting evidence in the data not considered. Little systematic attempt to consider evidence in terms of suggested answers. Tentative attempts to weigh up information—attempt restricted to only one compensation—usually unsuccessful leaving the problem unresolved.	
	Occasional tentativeness about answer yet why the feeling should exist is not explained.	
IV Explainer responses. Careful, systematic collection and analysis of data.	A. Successfully weighed up data—setting one piece against another—compensating one piece against another. Must have two compensations. Ruled out alternative solutions.	7
	B. More information used. More processing and more information drawn from outside the given data. Greater attention to the clarification of the problem. All alternatives were examined systematically. Was aware of gaps in information.	8
	C. Major distinction was quantity of material. Almost all presented evidence was used. Conflicts were resolved.	9

Concept Tests

The concept tests were concerned with measuring the figurative schemes held by each child and the extent to which the training sessions influenced them. In an earlier discussion it was pointed out that concepts or figurative schemes arise from the individual's interaction with his environment. They are ideas in the mind of the individual that summarize, chunk or categorize many personal meanings attached to things or events. They represent a class of attributes or relationships which sometimes apply to specific concrete referents but on other occasions refer to an abstract class of things (Engelmann, 1969; Flavell, 1970; Imperatore, 1970; Denney, 1974; Sigel, 1974). Therefore, to measure the nature of the concepts held by individuals, tests had to identify the number and nature of the attributes the child associated with a particular word label. In this instance the particular concepts selected were poverty and city. Both concepts are part of many Social Studies programs used in primary schools. Both were included in the problems examined in the training sessions. Furthermore, two of the Problem Tasks used as criterion tests centred on them.

Sorting photographs was the particular technique used to find the dimensions of the child's figurative schemes. Taking the set of photographs associated with city, the child was asked to sort them into two groups; city and non-city. On completing the task he was asked what it was about the photographs in the city pile that made him put them there. A similar question was asked about the non-city pile. The responses were recorded on tape. Each photograph in each pile was then taken in turn. The child was asked to explain why he placed it in that particular pile. The same procedure was used with the poverty set. 'Put the photographs in two piles; one about poverty and one not about poverty.' 'Why did you put these photographs together?' 'What is it about this photograph that made you put it into the poverty pile?' Scores were based upon the number of attributes identified.

Photographs were used because they provided a relatively simple and consistent way of eliciting the dimensions or chunks associated with a child's figurative schemes. With each photograph the centre of attention was restricted. Because of this, less demanding search behaviour was required. Thus, the overall task was easier. In a sense the task was segmented by the use of sets of photographs. In each circumstance it was more likely that the child would reveal all the attributes he associated with a particular concept label. Also, the use of the photographs permitted the use of a consistent procedure from interview to interview. It was unnecessary for the interviewer to prompt and probe. Each photograph acted as a response facilitator.

Such consistency in administration was essential in a test designed to measure the influence of training between experimental groups.

The photographs were selected to represent the various facets of cities and poverty. Twenty-one photographs were included in the *City* set. The *Poverty* set contained 16 photographs.

Finding Answers

This test was designed to assess the ability of the student to generate answers to given problems. There were ten problems in all, each focused upon a social dilemma. They included pollution, road safety, famine, and infectious disease. Each problem was described in a few short sentences. The student was expected to respond to the task by inferring answers from the information provided, by relating the problem to previous experience or simply by guessing. Performance was scored on the basis of number of possible answers suggested. No attempt was made to assess the quality of the responses.

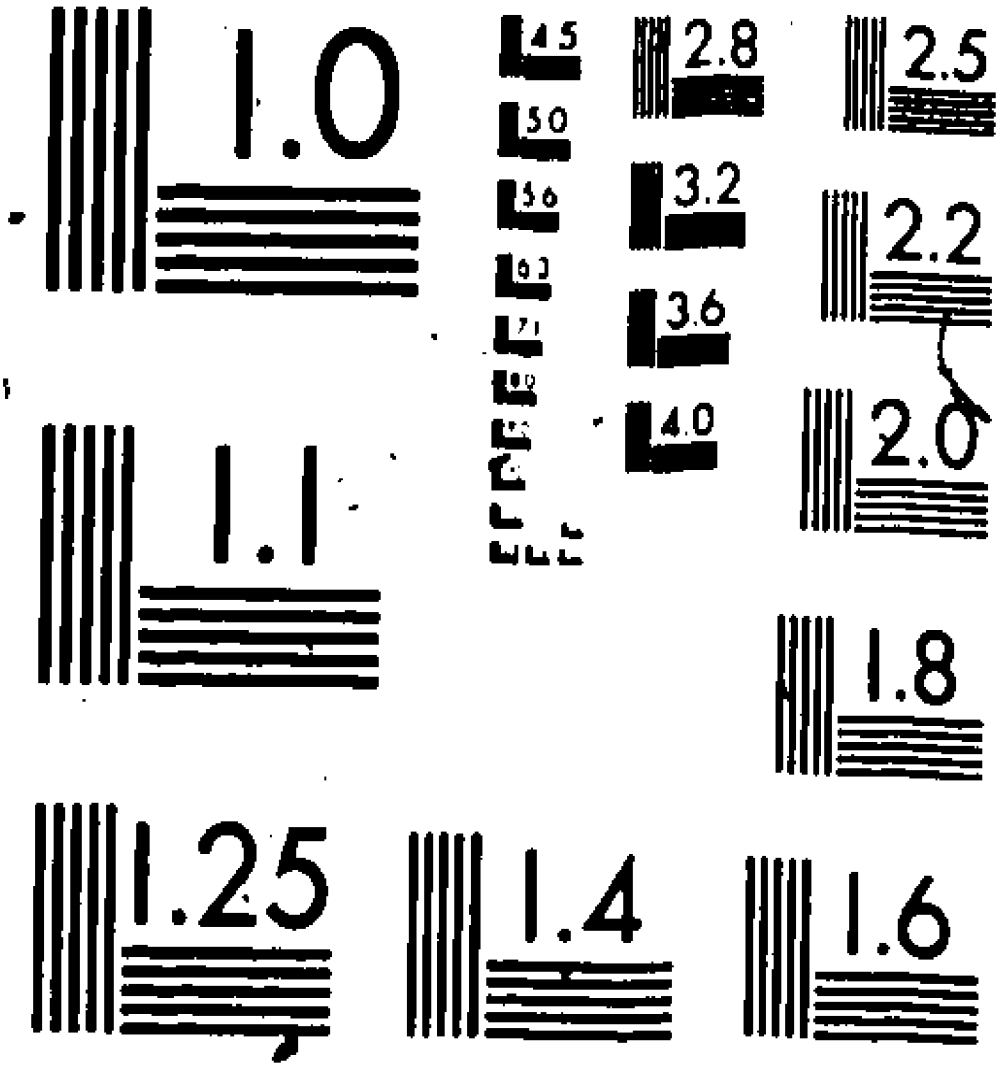
The test was administered in a group situation. After a practice item, students were taken through the test in a lock step fashion by the teacher. Students followed the information with their eyes while the teacher read it aloud. When all students had completed their response to an item, they were directed to turn to the next question in the test sequence.

The procedure tested with a group of 48 grade six students revealed no difficulties. Analysis of the results of this trial indicated strong internal consistency between the ten situations presented. An alpha coefficient of 0.88 was recorded.

Puzzles

This test, made up of 10 items, aimed at discovering the ability of the child to find evidence to support a given problem solution. In all instances the problems related to situations portrayed in photographs. A photograph was given, a problem posed and a solution suggested. The subject had to record as many reasons for the given answer as he could find. Each piece of evidence listed was given a score of one.

Administered in a lock step fashion, in an identical manner to that of the *Finding Answers Test*, the procedure proved to be relatively simple. Respondents were introduced to John. His photograph was on the first page of the test. John, they were told, had already answered all the questions. They had to find the reasons why he gave the answers he did. By adopting this procedure the nature of the task was considerably restricted, thus allowing attention to focus on a particular skill. The operation of finding evidence in photographs was one that had received considerable attention in the training sessions.



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

With one exception the photographs were of situations outside the culture of the respondents but of a type commonly found in Social Studies text books. Amongst the 10 photographs was one showing men building a hut in Kenya, another showed a young Buddhist monk working in his room, while another showed a young woman cooking a mea on a street in Hong Kong. The questions associated with the photographs were relatively simple. 'Is this woman at home or at work?' 'It this boy an Australian?' 'Are these people likely to stay long in the one place?' In all cases, the photographs provided a mass of information that could have been used to support the answer given by John.

While this form of test was not one experienced by students before, it was nevertheless well received. Formally tried out with 48 grade six students, the test proved to have good internal consistency between the items. An alpha coefficient of 0.81 was recorded.

Working it Out

Designed to assess the Student's ability to manipulate or weigh up evidence, this test consisted of seven social dilemmas of varying complexity. The situations included where to shop, which bicycle to buy, who should be elected team captain, what form of public transport should be provided for a new town, and where a new factory should be located. For each problem, two or three alternative answers were provided. Under each solution, relevant evidence was also presented. It varied from two to four points. Up to 12 pieces of data were given, although more could have been inferred. This arrangement allowed the subject to demonstrate his ability to manipulate and evaluate information in a careful, complete and systematic manner.

This test, like the previous three, related directly to Treatment One in the training program described above. In that treatment, a record sheet was used which made use of boxes in a similar fashion to that used in the *Working it Out Test*. In the training sessions, the student had to identify his own solutions, collect information, and make a decision. *Working it Out* focused on the final element of those three components of enquiry.

Performance was rated on a seven-point scale according to the amount of information related together, the extent to which negative information was included, the attempts made to modify a given solution to encompass some of the advantages listed under alternative answers, and a willingness to see the need for tentativeness due to deficiencies in the given information. The levels of performance reflected increasing sophistication in data processing.

Administered in a group situation, no difficulties were encountered in the trials. The subject proceeded from problem to problem in unison under the

direction of the test administrator. A try-out of the test with 48 subjects indicated that the test had a high level of internal consistency. An alpha coefficient of 0.85 was recorded.

Higgins Inference Test

Entitled the *Production of Inferences Task*, the test consisted of six parts each requiring a separate administrative session. The subject was required to draw inferences from a photograph for a period of 20 minutes. The number of inferences that were consistent with the given information constituted the score for each photograph. Higgins suggested that four rather than six photographs could be used. The reliability of the measures varied according to the number of parts used. With six photographs Higgins (1974) reported a reliability coefficient of 0.90; with four, the coefficient was 0.84.

Other Tests

ACER Intermediate D

Developed and published by the Australian Council for Educational Research, this test was described as a general ability or intelligence test. It utilized material of a verbal and numerical nature. Seventy-five items were arranged in ascending order of difficulty. The items included analogies, classifications, synonyms, number and letter series and questions involving arithmetical and verbal reasoning. The emphasis in the test was heavily verbal. Only 24 out of the 75 items concerned number and of these 13 were mathematical problems couched in words. Reported reliability coefficients, both split half and test-retest, were high (0.94 and 0.92 respectively).

ACER Word Knowledge Test Form B

In this test, the subject had to identify the synonym of a given word. Each item had an underlined word on the left and the child had to choose the closest synonym from the five words listed on the right. The 100 words listed were drawn from a general rather than a specialized vocabulary. A split half reliability coefficient of 0.92 was reported.

Backward Digits from WISC

This test measured the immediate memory span of the subject. It called for the repetition of a set of digits backwards. All items were given orally with each digit in the item presented at the rate of one per second. Each item had one alternative set of numbers in the event that the first set was missed. All items were placed according to increasing serial numbers starting with two digits backwards and going to eight digits. A test re-test reliability coefficient of 0.75 was reported.

Group Embedded Figures Test

The test was made up of simple and complex figures. The outline of the simple figure was incorporated in the complex one but obscured perceptually by means of line and shading patterns. The simple figure had lost its identity as a separate perceptual unit. The child had to trace the outline of the simple figure. Eighteen items were used.

Test of Words used in Social Studies, Form Y

Developed by ACER from a list of Social Studies words which teachers thought children should know, this test contained 38 items. For each item the child was given a 'concept label' and asked to recognize either instances, or else a simple definition, of the presented term. Examples of words used include: employment, election, primitive and society. The child's response was selected from four alternatives.

Test of Comprehension

Designed to measure cognitive abilities associated with interpreting Social Studies materials, this test utilized a variety of stimulus material. Maps, photographs, graphs and statistical tables were used. The abilities being assessed included understanding of the explicit and implicit meaning of the information given, and deducing consequences or associations. The test consisted of blocks of multiple choice items based upon separate pieces of material. In all, 52 items were provided at the grade six level.

Issues to be Investigated

The major questions of this investigation all relate to the effects of training upon problem-solving performance. An examination of Pascual-Leone's theory of cognitive development suggested that problem-solving training should direct attention at developing and enhancing appropriate executive, figurative and operative schemes. Subsequently, two forms of instruction were developed. One treatment consisted of intense instruction; where the appropriate skills were explained by the teacher and practised by the subjects. The second group discussed the same problems but received no explicit instruction from the teachers on the nature of the tasks or how they might set about solving them. Thus, on the criterion measure (Problem Tasks), it was expected that Treatment One, the intense training group, would perform significantly better than Treatment Two. Such a situation was expected because of the short duration of the training program. Given a much longer period, say 12 months, of problem-solving activities, Treatment Two may perform in a similar manner to Treatment One. However,

without teacher assistance, six weeks seemed inadequate to identify and develop the appropriate mental schemes. Both groups were expected to perform significantly better than the subjects in the control group: Treatment Three. These children had no experience with the Social Studies problems.

To examine more closely the influence of the training sessions a number of measures were employed as 'measures of take'. They were used to answer the question 'Did the training sessions enhance the figurative and operative schemes of the children involved?' These instruments were grouped together under the title, 'Element Tests'. The first set, Concept Tests, were designed to measure two figurative schemes, city and poverty. Both concepts were included in the training sessions for both Treatment One and Treatment Two. The teaching techniques were identical for both groups. Therefore, no differences between Treatment One and Two were expected. However, as Treatment Three did not receive any training its performance was expected to differ from both Treatments One and Two.

The Element Tests—*Answers*, *Puzzles* and *Working it Out*—were designed to measure particular operative schemes. Subjects in Treatment One received systematic instruction in these cognitive skills, whereas those in Treatment Two did not. However, Treatment Two presented the opportunity to develop the relevant schemes as subjects discussed problems calling for the application of such skills. Thus, subjects in both Treatments One and Two were expected to perform significantly better than those subjects in Treatment Three where no such experiences were provided.

The *Higgins Tests*, also included in the Element Test category, concerned inference-making behaviour. Again, this skill was specifically drawn to the attention of subjects in Treatment One while those in Treatment Two would need to have deduced it for themselves. Those subjects in Treatment Three were not given the opportunity to identify or practise the skill. Therefore, it was expected that the performance of Treatment One would be significantly different from Treatment Two. Both these treatments would be significantly different from Treatment Three.

Other criterion measures included in the investigation were *Test of Words used in Social Studies* and *Test of Comprehension*. The first used some of the words associated with the problem discussed in the training sessions by both Treatments One and Two. Consequently, those subjects in Treatments One and Two were considered to have an advantage over those subjects in the control group who did not see the problems. Similarly with the Social Studies comprehension test. A number of items involved interpreting and drawing inferences from photographs. Other items used graphs, statistical

tables and maps. Some of the cognitive strategies used in the training sessions were relevant to some of these items. Therefore, it was expected that those children taking part in the training would perform significantly better than those in the control group. Again, differences between Treatments One and Two were expected because of the differences in intensity of the training.

Two further aspects must be taken into account when judging the success or otherwise of the training experiences. 'Were the benefits of the training sessions durable?' and 'Did they transfer to similar although not identical type tasks?' Half the sample was re-tested two months after the completion of the post-testing sessions. Did any gains evident at the initial post-testing session survive to the delayed post-testing? Because subjects in Treatment One received a more intense and directed experience it was expected that their successes would be more durable. The Treatment Two experience was expected to be more susceptible to fading. Thus a real difference between Treatments One and Three was expected. Such was not the case with Treatments Two and Three.

To assist in examining the question of transfer, two forms of Problem Tasks were used. One employed photographic data and one used prose passages. Because the photographic material was almost identical to that used in the training sessions, the problem which was based upon it was not considered a transfer task. However, no experience was given with verbal material so the problem based upon a written passage was seen as a measure of transfer.

Because of the more intensive training involved in Treatment One, those subjects receiving it were expected to be in a better position to transfer their developed skills to a new task. Thus they were expected to perform significantly better than subjects in either Treatment Two or Treatment Three. Because Treatment Two had had some experience in problem-solving in Social Studies-type situations, the subjects in that group were expected to perform better than those in Treatment Three. Some experience was considered to be better than none.

Although not of critical concern in the present investigation, the influence of test-taking experience on subsequent performance was of interest. Did subjects who had had experience in a test situation with Problem Tasks perform better on the next occasion than children who were facing such tasks for the first time? That is, did completing the tests constitute a learning experience?

One further issue of interest in the present research concerned the relationship between performance on component tests of enquiry and global mea-

sures of problem-solving ability. As noted in an earlier chapter, some training research summed performance across a number of component tests to gain a problem-solving score. Other research examined problem-solving ability by employing global measures. With the present study both forms of measures were used. While not critical to the main purpose of the investigation, it was of interest to examine the interrelationship between the global problem tasks and the component or element tests.

The present chapter has outlined the design of the investigation, described the sample; discussed criterion measures and discussed the expected relationships. By using a nested design, it was possible to partial out the influences of group membership and so identify the contribution of the two training programs. It is in the next chapter that the gathered data related to each issue is examined to determine whether the proposed relationships did in fact exist.

7. Results

To examine the questions raised in the previous chapter, multivariate analyses of variance were employed. The program used for these analyses of data was *Multivariance Version 5* (Finn, 1972). Finn (1974) pointed out that although multivariate analysis is not always applicable to specific problems, it is generally eminently suitable in educational research where multiple outcomes are being measured. It is particularly appropriate, where sets of measures, identified through procedures like factor analysis, have common components. However, Finn stressed that the set of measures used must also be conceptually meaningful.

It will be recalled that three sets of measures were employed as dependent variables in the present investigation—four Problem Tasks, seven Element Tests, and two Social Studies Tests. In the analyses described in the following sections the seven Element Tests were divided into three separate groups. Primarily this was done because different numbers of subjects had completed different groups of tests. Rather than dropping subjects to allow all analyses to be completed together, it was considered preferable to handle each group separately. Thus, in the concept test group there were two sets of measures—*Poverty* and *City*. These two tests were regarded conceptually as measures of figurative schemes held by the subjects. The three tests, *Answers*, *Puzzles* and *Working it Out*, formed the second group. Each of these tests involved components of the enquiry process, thus constituting a meaningful group. In Pascual-Leone's terms they were measures of particular operative schemes. The *Higgins Inference Tests* also involved a component of the enquiry process but were handled separately because more subjects had completed them. The Australian Council for Educational Research Social Studies Tests were more general in nature, assessing a number

of comprehension skills and terms associated with Social Studies objectives, thereby forming a group of their own.

An inspection of the principal components analysis of the data indicated that each group of measures shared a high level of common variance. Such results supported the decision to use multivariate analysis.

In addition to the multivariate analysis, the Finn program provides for each dependent variable a univariate analysis of variance. Finn (1974) advised that with separate analysis of each measure, statistical error rates may be multiplied many times and the reproducibility of the study is reduced. A multivariate analysis focuses on the entire response after giving appropriate consideration to the correlations among the outcome measures. It reflects variation in the data as a whole, not only the separate parts. However, when the univariate analysis is used in conjunction with multivariate analysis the interpretation of results is assisted.

Analyses of Results

The results of each set of questions discussed in the previous chapter have been presented in turn. In each case consideration has been given first to the multivariate data. Decisions to accept or reject the given hypothesis rested upon the results of the multivariate analysis although the univariate data was inspected to identify the measures contributing to the significance. However, to discover the source and direction of the differences, estimates of effects have been examined. The appropriate tables for each question are included in the appendix.

While each analysis was designed to investigate a particular question concerned with treatment effects, comments have been made upon other significant effects revealed. Four factors were in fact considered: treatment, teacher, school and class. All were regarded as fixed effects. Classes were nested in schools, and schools nested within teachers. Although the source of the differences can often be identified, the reasons for such differences are not always clear. However, it has been important to acknowledge their existence because such differences provide areas for further research.

Did Performance of Treatment Groups Differ on Problem Tasks?

Here the issue was whether the experimental treatments influenced performance on four Problem Tasks—*Poverty*, *City*, *Indians* and *Famine*. The sole significant multivariate test statistic was concerned with treatment. It

was significant at the .001 level. Thus, a statistical difference between treatment groups was accepted. Inspection of the univariate analyses supports the multivariate conclusion. On the four Problem Tasks, all univariate statistics were significant.

The treatment effects were in the predicted direction. The mean performance of Treatment One was superior to the mean performance of Treatment Two which, in turn, was superior to the mean performance of Treatment Three, and the mean differences were significant on all four Problem Tasks.

Single significant univariate F ratios were indicated for both *Indians* and *Famine*. The first concerned a teacher effect while the second was an interaction between Treatment and Class. As the decision was made to accept or reject a statistically significant relationship on the basis of the multivariate statistics only, the source of these effects has not been discussed.

Did Performance of Treatment Groups Differ on the Concept Tests?

One of the contentions of the research was that the teaching programs would enhance figurative schemes. In particular the concepts *Poverty* and *City* were placed under separate scrutiny. The anticipated difference between treatment groups was evident at the .001 level of significance. Teacher effects were also significant but at the .05 level of probability. The univariate statistics for both measures supported the multivariate data on treatment effects, although the source of the difference with teacher effects appeared to be derived from the one test—*City*.

Treatment Effect

The performance of both Treatments One and Two were significantly different from the performance of Treatment Three, the control group, on both the *Poverty* and *City* tests.

Other Effects

The teacher effect revealed in the multivariate analysis appeared to be located in the *City* instrument. The univariate analysis indicated an F ratio of significance at the .05 level. Although noted, the nature of the teacher effect seen in both the multivariate and univariate analyses was not identified. The nominated contrasts applied (by computer) in the analyses were chosen specifically because of certain identifiable relationships between treatments. The same relationships did not apply to teachers, but the corresponding contrasts were produced by the program in the case of significant teacher effects. As any enquiry into the effects of teachers on student performance was outside the scope of the present study, alternative methods of locating

the nature of the teacher effect were not pursued. The influence of teachers upon student performance was simply noted as an area requiring further research.

Did Performance of the Treatment Groups Differ on the Element Tests—Answers, Puzzles and Working it Out?

The third question concerned the influence of training upon scores of three Element Tests—*Answers*, *Puzzles* and *Working it Out*. Training was expected to improve performance on these three particular components of the enquiry process. Four multivariate statistics were significant. Treatment, teachers, and schools nested within teachers, were all significant. In addition the interaction Treatment by Teacher proved significant.

Although the effects of Treatments, Teachers, and Schools were clearly significant, both Treatment and Teacher were confounded by the significant interaction effect between Treatment by Teachers. This interaction confounded the main effects in two ways. First, although the test of interaction was made eliminating main effects, main effect tests were confounded with interaction sums of products. Secondly, for interpretation, the existence of interaction suggested that tests of main effects may not have been valid (Finn, 1974). However, in the present case the main effects of Treatment and Teacher were so substantial and so much more powerful than the interaction effects, that it could be assumed there was something of consequence to report and discuss.

Treatment Effect

The multivariate statistic indicated a highly significant difference between the performance of the treatment groups. From the univariate analysis, the major source of this difference was performance on the *Working it Out Test*, the instrument concerned with the ability of students to manipulate and evaluate data. However, a significant treatment effect was also identified in the *Puzzles Test*, the test where children were asked to find evidence for a particular problem solution.

The mean differences in performance between the Treatment One group and the Treatment Three group, on the *Working it Out Test*, was accepted as being statistically significant. The contrast between Treatment Two and Treatment Three was not significant. Performance differences between the subjects of Treatment Two and the subjects of Treatment Three as reflected in mean scores were viewed as chance sample fluctuations.

On the *Puzzles Test* subjects in Treatment Three, the control group performed better on average than subjects in Treatment Two who had been

given opportunities to consider and collect evidence to support proposed problem solutions. Such a result appeared meaningless. However, the ratio of estimates of effects to standard error of estimation failed to reach an appropriate critical level so the difference between Treatments Two and Three were seen as a chance fluctuation in the sample – a more plausible situation, considering the experience of the groups.

Teacher Effect

The significant multivariate statistic for teacher effect was supported by three significant univariate F statistics. For the *Answers* and *Puzzles Tests* the students working with both Teacher One and Teacher Two performed significantly better than those working with Teacher Three. With the *Working it Out Test*, the Teacher One group was significantly different from the Teacher Three. The differences in performance between the Teacher Two group and the Teacher Three group on *Working it Out* was accepted as a chance difference.

There appeared to be two possible explanations for the teacher effect. First, the teachers were not equally successful in teaching the skills measured on the three Element Tests. Secondly, the teacher groups were not equivalent prior to the commencement of the experimental programs. Therefore, the observed differences at the conclusion of the research simply reflected the lack of initial group equivalence.

While differences in teaching styles between teachers were noted in the analysis of tapes made of three teaching sessions, it was also observed that the content covered by all three teachers was identical. Pre-testing of half the sample on four Problem Tasks did not reveal any significant teacher group differences. However, results on a general ability measure and a vocabulary test did indicate significant differences between teacher groups. The direction of these differences was identical to that of the differences noted on the three Element Tests. Yet it is intriguing that differences between teacher groups were not consistent across all measures. While the observed differences between the performance of the Teacher Two and Teacher Three groups were accepted as being significant differences on two measures, on the third they were considered to be due to chance. If a group was disadvantaged on one component of enquiry, why was it not similarly disadvantaged on other measures? What were the characteristics of these two teacher groups that accounted for the significant differences in performance on the three measures? Such questions require further research.

School Effect

The multivariate statistic for Schools within teacher effect was significant at the .001 level of probability. The univariate analyses on two measures

supported the conclusion of the multivariate analysis, namely, that there were significant differences between schools nested within teachers. However, the only significant contrast related to the comparison between Schools Three and Four. This was the first analysis in which a significant schools effect was revealed. Both schools were situated in the same municipality and, as far as could be determined by examining school registers listing parental occupations, the schools drew from a similar type of home background. Size of school and facilities available were equivalent. Moreover, the nature of the programs provided in the two schools appeared to be the same. Obviously, a deeper analysis of the school, its objectives and the programs provided would be required to identify the reasons for the existing school differences in performance on the two measures *Answers* and *Working it Out*. Perhaps the differences were attributable to specific prior learning. Alternatively, it was possible that students in one school compared to another did not put maximum effort into completing the two tests. Further research is required.

Treatment Interaction Effects

The significant interaction between Treatment and Teacher reported from the multivariate analysis was supported by one univariate F statistic gained on the *Working It Out Test*. Such a result suggested that some teachers operated relatively better with one treatment in comparison to another. All teachers did not handle all methods equally well. Teacher One performed significantly better with Treatment Two compared to Treatment Three, than did Teacher Three. Teacher Two was less effective with Treatment One compared to Treatment Three, than was Teacher Three. This interaction between Treatment and Teacher may be explained tentatively in terms of teacher style. Comment was made earlier on the differences between the three experimental teachers in their methods of discipline and the ways they encouraged their students to greater effort. It was likely that some teaching styles more comfortably accommodated certain teaching procedures. Treatment One cast the teacher in a dominant role, where she determined the direction and nature of discussion. With Treatment Two there was a more open approach, where the children were permitted to control the direction of their own enquiry, where the teacher probed and aimed to capitalize upon the student's own view of the problem and the skills needed to reach a solution. In a sense, the second treatment employed a guided discovery approach.

The idea that teacher style was a factor in explaining the teacher by treatment interaction appeared to accord with information about Teacher One. This teacher, by her actions, seemed to favour a less dominant role than

that demanded by Treatment One. Her style appeared to fit more naturally into the Treatment Two approach. Teacher Two, on the other hand, was a resolute classroom operator with a forceful personality. It therefore seemed plausible that her teaching style would match better the Treatment One approach than Teacher Three whose style was less vigorous. Yet examination of the data indicated that this *post hoc* prediction did not happen. Comparing achievement of Treatments One and Three, Teacher Two was significantly less sucessful than Teacher Three. Perhaps the style of Teacher Two was so intense that students in Treatment One reacted against it.

The comments above on the relationship between teacher style and treatment strategies must be viewed as being highly conjectural, particularly as the significant interaction between Teacher and Treatment arose only on one measure. If the *post hoc* view regarding teacher compatibility with treatment was correct, similar interactions might have been expected on other criterion measures. Previous analyses indicated no such interactions. Obviously the relationship between teacher style and advocated teaching strategies in curriculum programs needs further research.

Did Performance of Treatment Groups Differ on the Higgins Inference Tests?

The *Higgins Inference Tests* were a set of measures associated with an enquiry skill. Question four concerned the influence of the training program upon inferencing behaviours as measured by these instruments.

Treatment Effects

There proved to be no significant difference between the performance of the treatment groups. The multivariate statistic did not reach an acceptable level for rejecting the view that there was no difference. The univariate F statistics for both tests, *Higgins Two* and *Higgins Four*, supported the multivariate decision. Neither test had an F ratio which reached the .05 level of significance.

Other Effects

Significant differences between Teacher effects and Schools within teacher effects were noted. The probability of these differences being due to chance fluctuations in the sample was rejected at the .001 level of significance. Support for the significance of the multivariate Teacher and School effects was seen in the univariate analyses.

On both *Higgins Tests* the performance of the Teacher One group was significantly different from the performance of the group working with Teacher Three. Similarly with the Teacher Two group, whose performance was significantly different from the Teacher Three group.

The nature of the Teacher effects observed on the *Higgins Tests* was consistent with performance of the respective teacher groups on the *Answers* and *Puzzles Tests*. These latter tests were also examining student performance on components of the enquiry process. In discussing the results of the *Answers* and *Puzzles Tests*, possible explanations for the significant Teacher effects were seen to be varying teacher effectiveness or initial group differences in ability, prior to the commencement of the training program. Both factors could have been operating together. Both explanations, teacher effectiveness and initial group differences, were equally relevant in interpreting the results on the *Higgins Tests*. However, as discussed previously, no definite conclusion could be reached from the current research. Further exploration is necessary.

The source of the significant School effects related to the contrast School Three versus School Four on *Higgins Four*. The same contrast with *Higgins Two* almost reached the .05 level of significance. A significant difference between the performance of subjects drawn from Schools Three and Four was also observed on two Element Tests previously discussed—*Answers* and *Working it Out*. The difference between the two schools, in the current analysis, was in the same direction. As noted above, the reasons for such differences were unknown. On such attributes as school size, school facilities and curriculum objectives the schools appeared to be much the same. Obviously a detailed analysis of each school as a functioning unit was necessary if the observed performance differences of students were to be explained. The current research project did not include such an analysis. The relationship between school characteristics and student performance upon cognitive measures, such as the *Higgins Tests*, is an area requiring further research.

Did Performance of Treatment Groups Differ on the ACER Social Studies Tests?

Question five concerned the relationship between experiences with the enquiry process and performance on two Social Studies measures developed by the Australian Council for Educational Research. Designed to survey achievement levels of primary school children, these tests sampled a number of objectives. Compared to the Element Tests discussed previously, they focused upon a wider range of skills and concepts.

Treatment Effects

The observed differences between treatment groups were accepted as being due to chance fluctuations in the sample. The multivariate statistic failed to

reach the .05 level of significance. Both univariate F tests supported the multivariate decision.

Other Effects

Significant differences were identified in the multivariate analysis on the School within teacher, and Class within school and teacher effects. School effects were derived from the Social Studies *Comprehension Test*. The contrasts School One versus School Two and School Three versus School Four were significant.

It will be recalled that the performance of School Three subjects was significantly different from School Four on *Answers, Working it Out* and *Higgins Four* tests. However, the *Comprehension Test* provided the first occasion where the contrast School One versus School Two was significant. Why students at School One should do significantly better on the *Comprehension Test* than School Two, and School Three subjects significantly better than those from School Four was unclear. Although explanations could be sought in terms of factors such as the socio-economic backgrounds of the children, school facilities, school programs and motivation of students, the available evidence did not reveal any substantial variation between schools on such attributes.

The multivariate F statistic for Class effect was significant at the .05 level. However, the univariate F results for the two Social Studies Tests did not support the multivariate decision. This lack of agreement between the multivariate and univariate results was not bizarre. Finn & Mattson (1974) pointed out that there is no necessary relationship between the multivariate test results and the separate univariate results. The multivariate test statistic reflects variation in the data as a whole, not only for separate pieces.

The source of the Class effect identified in the multivariate analysis was not explored, because the particular characteristics of classrooms which influence performance on given criterion measures was not an issue of concern in the present research. Appropriate exploration of such a relationship would require more precise measurement instruments than were used in the present investigation.

When interpreting the results associated with any analysis, the decision to accept or reject the existence of significant differences between groups on a given factor rested with the multivariate F test. With the Teacher effect, two significant univariate results were reported, but the multivariate test failed to reach the .05 level of significance. The test statistic was, however, only slightly less than the critical value.

Were Differences between Treatment Groups Durable?

An important aspect of the experimental design was a delayed post-testing session involving four measures of problem-solving ability. This facet of the research was planned to examine the durability of any training effects initially noted at the post-testing session. The results of this delayed post-test analysis indicated significant multivariate tests for Treatment and Teacher effects together with a significant Treatment by School interaction.

Interaction Effects

The interaction of Treatment by School nested within teacher, significant in the multivariate analysis, was supported by two significant univariate analyses. These results suggested that the durability of training effects was not consistent between schools and treatments. Some treatments were more durable in some schools than others. On the *Poverty* measure two contrasts reached an appropriate critical value for the *t* distribution to accept the differences between means as being statistically significant, rather than being due to chance fluctuations. These contrasts concerned Treatments Two and Three with Schools One and Two, and Treatments Two and Three with Schools Three and Four. By comparing the Treatment Two versus Treatment Three contrast with School One versus School Two, it was seen that the training effect of Treatment Two was durable in one school but not the other. Training effect, as measured on the test *Poverty*, while durable in School Two appeared to have faded in School One. A similar situation pertained to Schools Three and Four. There the training effect of Treatment Two, when compared to Treatment Three, the control group, persisted in School Three but was lost in School Four.

For the *Goldfields* measure two contrasts were significant. One concerned the relationship between Treatments One versus Three with Schools One versus Two, while the second significant contrast referred to Treatment Two versus Treatment Three with Schools Three versus Four. For school One the durability of Treatment One compared to Treatment Three was significantly different from School Two. In School One the subjects of Treatment One continued to out-perform subjects in Treatment Three. Such was not the case in School Two. In School Two, the control group mean was superior to the mean performance of subjects in Treatment One. Therefore, in School Two the training effect of Treatment One was not durable.

With the significant contrast involving Schools Three and Four, the Treatments concerned were Two and Three. The expected differences in favour of Treatment Two when compared to Treatment Three continued to

exist in School Three. However, that was not the case with School Four. In School Four, subjects from the Control group, Treatment Three, performed better on average than subjects from Treatment Two. For Treatment Two, the training effect as measured on *Goldfields* was durable in School Three but not in School Four.

Why training effects should persist in some schools and not in others was not apparent. There appeared to be no significant differences between schools within each set: School One and School Two, and School Three and School Four. For example, all were built at approximately the same time, all had similar educational facilities, all drew pupils from similar socio-economic areas and provided what appeared to be similar educational programs. It was possible that after the training program was completed some schools provided experiences which complemented the original training. Yet the lack of consistent patterns in the results marred this explanation. Only in School Three were durability effects on the one treatment noted on both measuring instruments. School One achieved durability of training on one measure for Treatment One but not on the second measure. School Two with Treatment Two achieved durability on one measure but not on the second. Did the form of the test—visual versus verbal—influence the results? Obviously satisfactory explanations for the results gained were not evident in the information available to the investigator. Further research employing appropriate controls and measures would be necessary to tease out the particular factors accounting for the interaction between School and Treatment.

In addition to the significant interaction between Treatment and School, there were two significant main effects. One was Treatment and the second was Teacher. Because of the nature of the analysis, the Treatment by School interaction confounded the Treatment main effect, and interpretation of the Treatment effect must therefore be undertaken with caution. As noted previously when discussing the interaction concerned with the Element Tests—*Answers, Puzzles* and *Working it Out*—main effect tests are confounded with interaction sums of products.

Treatment Effect

Because the treatment effect, as reported in the multivariate data, was not significantly stronger than the interaction effect, and because support for the significant multivariate effect of Treatment was only found in one univariate analysis, it was considered more appropriate to accept the view that there were no statistical differences between the performance of the treatment

groups on the delayed post-testing measures. Therefore, the training effects noted at the conclusion of the training sessions were not seen as being durable.

Other Effects

The significant Teacher main effect on the multivariate analysis was supported by a significant F statistic for the univariate analysis associated with the *Goldfield* measure. Differences between teachers were significant at the .001 level of probability. The contrast between Teacher One and Teacher Three was significant. A similar significant difference between Teacher One and Teacher Three was identified in earlier analyses. On those occasions two possible explanations for the difference were explored. Although no firm conclusion was reached, consideration was given to the existence of initial ability differences between teacher groups prior to the commencement of training, and the possibility that teachers were not equivalent in terms of teaching competence. Both were applicable in the present analysis. Both were plausible explanations. Further research, however, would be necessary to resolve the issue.

Did the Performance of the Treatment Groups Differ significantly on the Transfer Tasks?

The seventh question concerned the transfer effects of the training program. On question one it was accepted that on four Problem Tasks the observed mean differences between treatment groups were statistically significant variations, rather than chance sampling fluctuations. But did the effects of the treatments provided only relate to tasks bearing a strong similarity to those undertaken in class, or did the effects also carry across to tasks employing a different data form? The univariate F statistics for the *Indian* and *Famine* tasks indicated statistically significant differences in levels of performance between the treatment groups, thus suggesting that transfer of training did occur. However, the question concerning transfer could not be resolved by inspecting the univariate analyses. Finn & Mattsson (1974) warned that interpreting the separate univariate results is likely to inflate statistical error rates dramatically. Moreover, where criterion measures have non-zero intercorrelations the univariate F statistics are not independent of one another. The performance of subjects on the measures *Indians* and *Famine* were related as the earlier discussion showed. Therefore, a multivariate analysis, using the two verbal Problem Tasks as criterion measures was necessary.

The results of a multivariate analysis showed two significant effects upon performance. One was a treatment effect and the other a teacher effect. The univariate F statistics for both Problem Tasks concerning the treatment effect, supported the conclusion derived from the multivariate analysis. The source of the difference for teacher effects was with the one measure—*Indians*.

Treatment Effect

Both treatment contrasts, Treatment One versus Treatment Three and Treatment Two versus Treatment Three, were significant on both measures—*Indians* and *Famine*. Therefore, it was accepted that the benefits derived from the training sessions did transfer to Problem Tasks utilizing a different data form.

Other Effects

From the univariate analysis it was seen that the source of the teacher effect identified in the multivariate analysis was primarily located in the *Indian* test. On that test both contrasts reached the critical *t* value in a two-tailed test of significance at the .05 level. The contrasts between Teachers One and Three and Teachers Two and Three were significant. Similar significant differences between the teacher groups were found in other analyses using different measures. On those occasions when possible explanations for the relationships were explored, no firm conclusions were reached. The same explanations are applicable here, namely, inequality between teacher groups prior to the commencement of training and variation in teacher effectiveness.

Did the Experience of doing the Tests on one occasion significantly influence Performance on subsequent re-testing?

In addition to the major questions of the current research, there were two subordinate issues which were of interest. One concerned the influence of the first administration of the tests upon performance at subsequent administrations. The second area of interest concerned the relationship between performance on the Problem Tasks and performance on tests measuring separate abilities of the problem-solving process.

To examine the first issue two analyses were undertaken. It will be recalled from the discussion on the design of the investigation that a control group, known as Treatment Three, was formed at the beginning of the project. Subjects in that group received no special training on problem-solving. From that group half of the subjects were both pre- and post-tested. The other half only com[illegible]d the Problem Tasks in the post-testing session.

By comparing the performance of these two sub-groups on the post-test tasks, the degree to which a test-taking experience influenced subsequent performance could be identified. The results of this analysis indicated that the multivariate F statistic was not significant. Support for the multivariate decision was seen in all univariate analyses. Any variation in performance at the post-testing session between those subjects pre-tested compared to those not pre-tested was therefore interpreted as due to chance alone. The pre-test experience on the Problem Tasks did not significantly influence later performance.

A second analysis using a different set of data was carried out to examine the same issue. It will be recalled that half of the subjects from all three treatment groups were pre-tested to check the equivalence of the groups prior to training. Further information regarding the extent to which the test-taking activity was a learning experience, was gained by examining the performance of the pre- and post-tested groups on the Problem Tasks compared to the post-test only group on the same Tasks. Note that in this analysis some subjects had training sessions intervening between the pre- and post-administration of the tests while others had engaged in training sessions prior to the testing session. It was therefore possible to discover an interaction between pre- and post-testing versus post-testing with treatment group. Thus, in the analysis, provision was made to identify such an interaction should it have occurred. However, a significant interaction did not occur, nor was there a significant difference between the pre-and post-test group and the post-test only group. This conclusion applied to all the univariate analyses as well as the multivariate analysis.

To what Extent was Performance on the Problem Tasks related to Performance on the Element Tests?

In many projects examining the impact of training upon problem-solving performance, the researchers utilized tests which measured components of the enquiry process. Although performance across these separate tests was often summed, the suitability of such instruments as measures of performance on the total enquiry task was questioned (see Chapter 4). As the present investigation used tests which measured elements or components of enquiry together with measures reflecting more closely the total process, it provided an opportunity to examine the relationship between these two broad groups of tests.

To explore the relationship between the two sets of tests in linear combination the multivariate technique of canonical correlation was employed

(Cooley & Lohnes 1971). Darlington, Weinberg & Walberg (1973) indicated that the technique was suitable for the proposed purpose.

When all the variance on the Problem Tasks was considered, only 15 per cent could be predicted from all four factors from the predictor battery. It was clear, therefore, that while there was a significant canonical function linking the two sets of measures, the majority of the variance associated with the major factor of each battery was unique to that battery.

Two alternative interpretations can be placed upon the result of the canonical analysis. First, the component tests were incomplete in their sampling of the abilities associated with the total process of enquiry. Perhaps some key abilities in the process were not included in the test battery. Secondly, irrespective of the number of component tests used, there is a difference of significance between a performance score on a task which requires marshalling, co-ordinating and applying many skills, compared to a performance measure derived from test scores on a number of separate enquiry abilities. If this second conclusion were correct, then research which utilized only component skill tests to evaluate the success of an enquiry training program placed serious restrictions upon the applicability of the findings regarding the suitability of the training program for influencing the problem-solving performance of students.

Summary

Nine issues were advanced for investigation. Seven related to the effects of two training programs upon performance of students on a number of criterion measures. These measures were seen to fall into three categories. The first and most important group concerned the problem-solving tasks. The second group concerned performance on specific abilities involved in the enquiry process. The third group were multiple choice Social Studies tests.

In an effort to isolate the specific contribution of the treatments to performance, an experimental design was adopted which allowed effects of teacher, school and class to be separated. This gave an opportunity to examine the way these factors influenced performance. Such observations, however, were seen as providing insights into areas requiring further research rather than being of major concern.

Teacher Effect

Significant teacher effects were noted on five of the criterion measures. All tests concerned were measuring elements of the enquiry process—*Answers*,

Puzzles, *Working it Out*, *Higgins Two* and *Higgins Four*. In all but one instance, the performance of the subjects working with Teachers One and Two were significantly better than the subjects working with Teacher Three. The exception was the *Working it Out* test where only the Teacher One group performed better than the Teacher Three group.

Firm conclusion as to why such differences existed could not be established. The teacher group were initially significantly different from one another on measures of general ability and vocabulary. Differences between groups on those measures favoured the groups of Teachers One and Two compared to the Teacher Three group. Nevertheless, similar differences between the teacher groups did not appear on all criterion measures. If one teacher group was qualitatively different to the other groups, why was that difference not reflected on all-criterion measures? It seemed possible that some teachers were more effective than others in teaching certain cognitive tasks. From the transcripts of three tapes of lessons given by each teacher to treatment groups one and two, it was found that all teachers faithfully followed the planned lessons, but it was also observed that the teaching style of the three experimental teachers varied. Was it possible that some styles were more appropriate for the teaching of certain skills?

School Effect

School effects were noted on four criterion measures. The tests concerned were *Answers*, *Working it Out*, *Higgins Four* and Social Studies *Test of Comprehension*. On each occasion where school effects were noted the contrast between the performance of School Three and School Four was significant. On one occasion the contrast between School One and School Two was also found to be significant.

In terms of general ability and vocabulary scores, the schools within each of these two sets did not vary significantly from one another. On dimensions of school facilities, school program and socio-economic background of parents, there was little variation. Explanations for the identified differences on the criterion measures were not apparent. The school contrast where differences might have been expected, School Five versus School Six, did not prove significant on any criterion measure. School Six, drew from a slightly lower socio-economic group, had a higher migrant population than any other school, and performed significantly lower on the general ability and vocabulary measures than School Five. Yet School Six did not perform significantly differently from School Five on any criterion measure.

Interaction Effect

Two interactions were noted. In one instance the interaction was between Treatment and Teacher and in another between Treatment and School. The first set of interactions was on the *Working it Out Test* and the second set on the *Goldfields* and *Poverty* measures, Problem Tasks administered at the delayed post-testing session. For those particular criterion measures the results suggested that where treatments influenced behaviour, the benefits were not equally distributed between teacher groups nor between schools. Why the results of the treatments should need to be qualified in these ways was not clear.

Treatment Effect

The major intent of the project was to investigate the impact of two forms of problem-solving experience. To assist in this examination seven questions were proposed which related the expected benefits of the training programs to performance on a series of measures. Of the seven null hypotheses involved, four were rejected. Statistically significant differences were accepted between the performance of treatment groups on the Problem Tasks, the Concept Tests, the transfer tasks, and with certain reservations, the Element Tests. Observed differences between treatment groups on the *Higgins Tests*, the ACER *Social Studies Tests*, and the delayed post-test Problem Tasks were found to be due to chance. For each of these sets of measures the null hypotheses were not rejected. Discussion of the educational significance of these inferences made at a statistical level is presented in the next chapter.

8. Discussion of Treatment Results

Although research has indicated that primary school children have considerable difficulty in engaging in the enquiry process as a totality, enquiry has remained an important objective in modern Social Studies curricula. In an earlier chapter, the Pascual-Leone view of intellectual development was discussed, highlighting the constraints impinging upon cognitive performance. A teaching approach which could prove beneficial in introducing the student to the enquiry process was also identified.

Basically, the constraints upon intellectual performance are three in number: limited available M-space or cognitive processing space; a tendency for some children to seize upon perceptually salient features in the data or a reluctance, on their part, to consider all the information provided; and lack of appropriate mental schemes—executive, figurative and operative. The training program adopted aimed at enhancing the schemes a child had while at the same time giving him strategies or procedures which, when used, would help to circumvent the other factors inhibiting his problem-solving performance. This program, entitled Treatment One, required the teacher to play the role of instructor. The teacher provided the students with information appropriate to the tasks and saw that they used it.

A second program (Treatment Two) was also developed. It contrasted with Treatment One in that the teacher did not provide direct instruction. The same problems were considered by both groups, but in Treatment Two the teacher gave no directions as to what was expected of the subjects, nor did she assist the subjects to develop appropriate mental schemes associated with problem-solving. Rather the students were expected to resolve for themselves the best method of approach and to identify the knowledge and

skills required. The teacher acted as chairman for the ensuing discussion. The overall approach in Treatment Two was apparently much closer to Piaget's view of what a teacher's role should be as the teacher neither explained nor demonstrated an appropriate course of action.

While there were differences between Treatments One and Two, there were also similarities. Subjects in both groups tackled the same problems. The initial phase of the enquiry was also treated identically with both groups. Children were instructed to list information from the photograph, to group items that shared common attributes, and to provide a label for each group that reflected the contents of the group. A number of Social Studies programs described this type of activity as a concept formation phase of a social study (Taba *et al.*, 1971; Queensland, Education Department, 1972; Victoria, Education Department, 1974). Through such activity the subjects became aware of and strengthened the figurative schemes associated with the given data and problem. Furthermore, the strategy reduced the possibility that some children would seize upon only the perceptually salient features or consider only a very limited amount of information because to do more required greater mental effort.

How Successful was Treatment One compared with Treatment Two?

Children in Treatment One were expected to out-perform children in Treatment Two. With Treatment One, appropriate problem-solving responses were detailed and instruction was given in a number of key skills. Similar assistance was not provided in Treatment Two. Consequently, the superiority of Treatment One over Treatment Two was expected to be reflected in performance on the Problem Tasks and some of the tests measuring components of the enquiry process. This section explores whether the expected relationship between Treatment One and Treatment Two eventuated.

Returning to those measures where significant treatment effects were noted, it can be seen from tables reporting estimates of effects that Treatment One made a more powerful impact upon performance than Treatment Two. (See Appendix.) For convenience, the relevant statistics between Treatments One and Two are represented in Table 8.

TABLE 8
Estimates of Treatment Effects in Standard Deviation Units

	Problem Tasks				Concept Tests	
	Poverty	*City*	*Indians*	*Famine*	*Poverty*	*City*
Treatment 1	0.80	0.87	0.83	0.90	0.70	0.65
Treatment 2	0.44	0.39	0.40	0.30	0.31	0.35

The estimates of treatment effects in standard deviation units for Treatment One were approximately twice those for Treatment Two on all Problem Tasks and Concept Tests. In most cases, when compared to the control group, the gains of Treatment One were twice as great as the gains of Treatment Two. For the Problem Tasks these results agreed with the predictions noted in Chapter 6. However, significant differences in performance on the Concept Tests between Treatment One subjects and Treatment Two subjects were not expected.

The training programs of Treatments One and Two shared some features (see Table 2). It was one of these shared attributes—the listing, grouping and labelling of information—that was expected to influence performance on the Concept Tests. Because the skills needed to improve performance on the Concept Tests were seen to be equally represented in the training programs of both treatment groups, any variation between the performance of Treatments One and Two on the Concept Tests was expected to be due to sampling fluctuations. However, the magnitude of the differences in estimated effects between the two programs suggested otherwise. In an attempt to explain this occurrence the training programs were re-examined. This *post hoc* analysis indicated that in addition to the listing, grouping and labelling activities there were other opportunities given in the programs which could have led to the refinement of figurative schemes. These opportunities, however, were not identical for both treatment groups. For example, with the problem associated with the photograph showing girls at a well in an Indian village, 'Do these girls live in poverty?', Treatment Two subjects proceeded to discuss the problem without teacher intervention. For Treatment One, the task was segmented into sub-tasks by the teacher. Subjects were encouraged to clarify the nature of the problem and to find as much evidence as possible to support alternative solutions. Such a procedure could have given an advantage to the Treatment One group, thus explaining the magnitude of estimated treatment effects existing between the two experimental programs.

Why was the Training Successful?

Highly significant differences between the mean performances of the treatment groups were found on the four Problem Task measures. As the mean performances of subjects in the two training groups were significantly better than the mean performance of the control, it was assumed that the former set of subjects were utilizing the limited M-space available to them in a more efficient manner than their classmates in the control group.

It will be recalled that in the theoretical model proposed by Pascual-Leone, size of M-space was seen as a major determinant in problem-solving performance (see Chapter 3). M-space was defined as the maximum number of mental schemes (mental blueprints which represent experience) that can be co-ordinated at any one time. The size of this information processing capacity was not amenable to training. M-space is largely hereditarily determined and grows as a function of age. It was argued, however, that it is possible to use the available M-space more efficiently by improving, developing and altering appropriate mental schemes. Three types of schemes were noted: figurative, operative and executive. Figurative schemes are the internal representations of items of information with which a subject is familiar, or of perceptual configurations which he can recognize. Operative schemes are the internal representations of functions, which can be applied to one set of figurative schemes in order to generate a new set. Executive schemes are the internal representations of procedures which can be applied in particular situations in order to reach particular objectives. As such, they are to a large extent responsible for determining what figurative and operative schemes a subject activates in any particular situation (Case, 1974*a*). The training programs aimed at establishing and enhancing all three types of schemes.

There was one further possible explanation for the more efficient use of M-space and that related to the strategies introduced in the training sessions—the listing, grouping and labelling of information gathered from the photographs. Such a systematic procedure may have assisted the field-dependent subjects to avoid the salient but irrelevant cues in the data. These subjects would then be filling their limited processing space with appropriate rather than irrelevant information, thus permitting better problem-solving performances.

The alternative explanations for the significant performance differences between the treatment groups on the Problem Tasks are examined below. To assist in this exploration, reference is made to the Concept Tests and Element

Tests—*Finding Answers, Puzzles, Working it Out* and the two *Higgins Inference Tests*. The Concept Tests were seen as measures of figurative schemes, whereas the Element Tests named were seen primarily as measures of operative schemes associated with the problem-solving process.

It will be remembered from the discussion in Chapter 3 that figurative, operative and executive aspects of knowing are integral to solving any given problem. In real situations figurative, operative and executive knowing never occur in isolation. The problem-solver, to be successful, must have information concerning the content of the problem, the skills to operate upon this information and a view of what he is trying to achieve. However, tasks can require different portions of these knowledge components. While success in problem-solving on one task may largely depend upon the availability of appropriate figurative schemes, for another task the presence of suitable operative schemes may be more crucial. The Concept Tests used in the present investigation were seen primarily as measures of figurative schemes because they aimed at identifying the attributes of the mental images a child associated with 'poverty' and 'city'. *Finding Answers, Puzzles* and *Working it Out* were considered to be primarily measures of operative schemes because they focused upon the ability of the child to operate upon given information, either to generate possible solutions to a problem, to find evidence to support answers, or to establish an appropriate solution.

If we assume that the Concept Tests and the Element Tests listed above were respectively measures of figurative and operative schemes, it then becomes possible to sort out the contribution made by enhanced operative, figurative or executive schemes to success on the Problem Tasks. If the success achieved by the two special treatment groups on the major criterion measures (the four Problem Tasks) were to be attributable to improved figurative and operative schemes, significant performance differences between these groups and the control would be expected on the Concept and Element Tests. Otherwise success on the Problem Tasks would have to be attributed to improved executive schemes. These possibilities are considered in the following discussion.

Improved Operative Schemes?

Only in Treatment One did subjects receive detailed instruction related to particular operative schemes. Attention was given to creating alternative answers to given problems, to gathering evidence to support alternative solutions, to manipulating evidence to reach a decision regarding a given problem, and to inferring information from details presented in photographs.

However, the subjects in Treatment Two did have the opportunity to develop such schemes on their own initiative as they, like the subjects in Treatment One, discussed problems where these mental abilities were applicable.

From the multivariate analysis of performance on the three measures—*Answers*, *Puzzles* and *Working it Out*—the null hypothesis expressing the lack of significant differences in performance between the treatment groups was rejected. This was done after acknowledging the existence of a significant interaction between teacher and treatment. Such an interaction confounded the treatment effect but because the latter was statistically so significant the difference in mean scores between treatment groups was not seen as a chance fluctuation in the sample. This result, at least initially, suggested that a source of the significant treatment effect as measured by the Problem Tasks was attributable to improved operative schemes. However, closer inspection revealed that the significant difference between treatments was present in only one of the three univariate analyses. Of the three measures, only *Working It Out* was associated with a significant treatment effect. But once again there was a confounding influence as the treatment by teacher interaction was statistically significant. If the differences between the performances of treatment groups on the Problem Tasks was to be attributed to improved operative schemes, the improvement had to be limited to the abilities associated with data manipulation. Differences in performance between the treatment groups on *Answers*, *Puzzles*, and the *Higgins Inference Tests* were not significant.

The view that the improvements in problem-solving performance arose from a refined operative scheme, however, was not satisfactory. If the establishment and enhancement of a particular operative scheme was the prime source of improved performance on a set of Problem Tasks then that source might reasonably be expected to apply equally well to explaining the success of both training programs. Previously it was noted that the performance of subjects in Treatment Two on the Problem Tasks, like those children in Treatment One, was significantly better than that of the subjects in the control group. However, Treatment Two subjects did not perform significantly differently from the control group on the *Working it Out Test*. The success of Treatment Two on the Problem Tasks therefore, could not be linked with an improved operative scheme. Consequently, the possibility that success on the Problem Tasks was due to an improved operative scheme was viewed with some scepticism.

Improved Figurative Schemes?

With the Concept Tests, *Poverty* and *City*, the mean scores of both training groups were significantly better than the control. On the Concept Tests also, the magnitude of the achievement differences between Treatments One and Three, and Treatments Two and Three, were similar to the size of the treatment effects between the same groups of subjects on the Problem Tasks. This was seen in Table 8. To the extent therefore that the *Poverty* and *City* tests measured figurative schemes, the explanation that the observed mean differences between treatment groups on the Problem Tasks was due to improved figurative schemes seemed plausible. With further analysis, however, this explanation lacked credibility.

Of the four Problem Tasks, two involved figurative schemes which were discussed in the training programs. The figurative schemes underlying the problems *Indians* and *Famine* were not discussed in class. It therefore followed that, if the aspect of training accounting for improved problem-solving performance was the attention given to figurative schemes, it would be discerned by comparing the performance of the treatment groups on the two sets of Problem Tasks—*Poverty* and *City*, and *Indians* and *Famine*. If figurative scheme improvement was the key factor in problem-solving achievement, the data in the study had to match two criteria. First, Treatment groups One and Two had to perform significantly better on *Poverty* and *City* than the control. Secondly, significant differences between the three experimental groups on the *Indians* and *Famine* should not exist. The results presented in the previous chapter indicated that the first requirement was met but not the second. Both were necessary if the differences in problem-solving performance were to be attributable to improved figurative schemes.

Further support for the dismissal of the view that superior performance of Treatment One and Two on the Problem Tasks was due to better figurative and operative schemes was seen in the canonical analysis. While there was a significant relationship between performance on some of the enquiry component measures—both figurative and operative aspects—and performance on the Problem Tasks, the relationship was described as moderate. Of the variance identified between the four Problem Tasks, only 15 per cent could be predicted from all four factors identified from the predictor battery, the enquiry component measures. The view that improved figurative schemes or improved operative schemes accounted for improved performance on the Problem Tasks seemed unsatisfactory.

Improved Executive Schemes?

The third major explanation for the observed treatment results concerned the development of appropriate executive schemes. Case (1972*a*) used the term executive scheme to encompass both what the student saw as task instructions and the nature of the problem task as a whole. In the present investigation it was assumed that, where a child had a clear view of what was expected of him in any task, he made more efficient use of his cognitive processing space. The competing but irrelevant schemes he may have considered, if he was unsure of what the task required, were not alerted. Thus the M-space that otherwise might have been occupied with irrelevant mental schemes was available to consider other possible answers and additional pertinent information. Because of this more intensive use, in terms of the given task, of available M-space, more sophisticated responses were made. In the terminology of Flavell & Wohwill (1969), through the provision of the special training, the performance of the students was brought closer to their true level of competence.

Of the explanations considered in explaining the success of the training program, the one concerned with better executive schemes appeared the most plausible. The variation in the magnitude of the treatment effects between the training groups could be related to the degree of explication in the instructions provided. For Treatment One, the teacher told the subjects the features of good answers and constantly urged them to include such features in their responses: clarification of the problem, consideration of alternative answers, collection of evidence and decision-making in terms of evidence gathered. For Treatment Two no such aid was provided, although the subjects had the opportunity to argue out their solution with their peers. In such discussions they were clarifying their own views as to what might be expected. If members of Treatment Two were to defend their decisions successfully against the arguments of their classmates, they had to consider all the information provided and to weigh up their solution against alternative explanations. Treatment Two subjects were setting up their own plan of action for handling Problem Tasks. Because the teacher did not specifically comment upon the adequacy of their procedures an element of doubt may have still existed, thus permitting the intrusion of less efficient strategies than those employed in Treatment One. Unlike the Treatment One activities, the success of the Treatment Two experience relied to a large extent upon the ability of the group members to debate the problems posed. Observations made by the experimental teachers suggested that, particularly in the early training sessions, the children involved had difficulty in considering the viewpoints of others and this limited debate.

One further possible explanation for the treatment effects should be noted. It was pointed out in earlier chapters that some children fail to use all available M-space either because they pay attention only to the perceptually salient features of the data and thus ignore a mass of other information that could be used, or because they are not stimulated by the task and therefore do not make the mental effort required to utilize all the M-space they have at their disposal. Such children are labelled field-dependent. The strategy of systematically scanning the given data in order to list, to group, and to label information was seen as a means of circumventing the disability suffered by field-dependent children. The use of this strategy could have been the basis for the improved problem-solving performance of the training groups. Unfortunately, the results did not support the explanation. As both Treatment One and Treatment Two were introduced to the same procedure, it would follow that, if the improved performance were due to the reduction of the constricting influence of field-dependency, then both treatment groups would have benefited equally and thus performed equally well on the Problem Tasks. As already indicated, this pattern in the results did not occur.

Saarni's (1973) research suggested that the construct of field-dependence was irrelevant for complex problem-solving tasks, although she conceded field-dependence might have a role in determining performance where problems were more perceptually bound. In particular, Saarni noted that within each Piagetian cognitive level field-independence did not affect the complex multi-step, problem-solving performance on the productive thinking problems used by Wardrop *et al.* (1969) and Olton & Crutchfield (1969). Yet Suchman (1961) observed perceptual biases in the search patterns of grade four students when presented with short motion films of simple physics demonstrations.

In the present investigation, two of the Problem Tasks shared at least a superficial similarity to those used in the *Productive Thinking Program* in that they utilized verbal materials. The remaining two Tasks used photographs. If Saarni's contentions regarding the relevance of field-dependence to problem-solving were correct, and, if the success of the training program in the present investigation were to be attributable to the assistance given to field-dependent subjects to adopt systematic scanning strategies, there should have been a difference in performance patterns between treatment groups on the verbal and visual Problem Tasks. Significant differences in student performance on the visual tasks would have been expected between Treatments One and Two, and Treatment Three as no assistance was given to

the latter group to overcome field-dependency constraints. Similar differences would not have been expected on the verbal problems because of Saarni's finding that field-dependency was irrelevant for such tasks. Such a pattern in the results did not emerge. There were significant differences in performance between treatments on all Problem Tasks irrespective of whether they used verbal or visual material.

A further reason for not attributing the significant improvement in problem-solving performance to the training program's strategy for desensitizing field-dependency was the lack of a strong relationship between performance on the Problem Tasks and the *Group Embedded Figures Test*, a measure of field-dependence. While two of the correlation coefficients appeared to be statistically significant, the relationship they reflected was only modest (0.5 and 0.15)*. It was of interest to note that both significant relationships related to the verbal tasks rather than the visual tasks. This result was not consistent with the Saarni (1973) conclusions.

In summary, of the various attributes built into the training program, the most plausible as an explanation for the treatment effects noted on the Problem Tasks appeared to be the identification or clarification of those executive schemes associated with the enquiry process. It was unlikely that improvement in figurative or operative schemes contributed in any major way to the success noted. Nor could the training effects be linked to aid given to field-dependent subjects.

Previous Research

No previous research in Social Studies enquiry training appeared to attribute problem-solving success to improved executive schemes, although Suchman (1961), Wardrop *et al.* (1969) and Robinson *et al.* (1972) saw such schemes as important contributors to performance. Suchman (1961) noted that it was the strangeness of the given tasks which caused the most difficulty. Early observations with the *Productive Thinking Program* convinced the authors that attention had to be given to what were termed 'master thinking skills'. Such skills were those associated with the deployment and co-ordination of a variety of enquiry abilities (Covington, n.d.). Yet, despite such observations, the major measurement thrust was in evaluating performance on separate skills. Attention, in the main, was not given to the ability

*This test of significance assumed a simple random sample, which was not employed. Significance would clearly disappear for the latter correlation coefficient if a correction were made for the design effect of the two stage sampling procedures used, while the former correlation would then be of borderline significance at best.

of the child to relate and co-ordinate a number of complex skills together in solving problems (Anderson, 1965; Butts & Jones, 1966; Hunkins & Shapiro, 1967; Olton & Crutchfield, 1969; Wallen *et al.*, 1969; Robinson *et al.*, 1972; Case & Fry, 1974; Higgins, 1974; Ojemann & Campbell, 1974). Outside the context of enquiry training, however, Kohnstamm (1970) claimed significant improvement in the performance of five-year-old children on a Piagetian class inclusion problem, because each child was given a detailed explanation of the task requirements.

Was There a Transfer of Training to Different Tasks?

Two of the problems given in the post-testing session were viewed as transfer tasks. Like those problems considered in training sessions, 'Where did the Indians obtain their horses?' and 'How can the shortage of food in India be overcome?', they were Social Studies-type problems which required the subjects to clarify the nature of the problem, suggest possible answers, collect evidence to support each possible answer and then consider this information to reach a decision. But neither of these two questions utilized information presented in photographic form. Both were based upon data presented in verbal passages. Thus, subjects had to transfer those understandings or schemes developed in training sessions to new situations embedded in a data form which was different from that originally considered in class. Were the subjects successful in making the transfer? The results reported in the previous chapter indicated that they were. Training on photographs was generalizable to verbal material.

The degree to which skills learned in one context transfer to another and the extent to which such skills are durable over time are two important criteria in assessing the contribution of a training program (Piaget, 1964; Inhelder & Sinclair, 1969; Snow, 1974). Social Studies curricula assume that abilities developed through the study of a particular issue will not be confined to that issue but that the child will be able to use skills to solve other problems. A few studies exploring the impact of enquiry training claimed significant transfer effects (Anderson, 1965; Olton & Crutchfield, 1969; Robinson *et al.*, 1972). The present study joined this group. Previous research, however, showed that particular enquiry skills or operative schemes transferred to new situations. Results of the present investigation suggested that subjects were able to transfer their overview of the enquiry process to new problems. That is, subjects were able to apply executive

[illegible] learned in one context to different problems set in a different [illegible]

Treffinger & Ripple (1970) were critical of earlier findings related to transfer. They pointed out that the instruments used bore too close a resemblance to the activities used in the training sessions. While it was true that the transfer tasks used were closer to the 'measure of take' end of the evaluation dimension identified by Robinson *et al.* (1972), they were still transfer tasks. Covington (n.d.) argued that where a task called for the application of thinking skills in topics with which the student had had little or no previous acquaintance, it was legitimate to view it as a transfer task. The criticism of Treffinger & Ripple applied equally to the present investigation because the transfer tasks bore some similarity to the problems used in training sessions. On the other hand the tasks where sufficiently different to be viewed as true transfer situations. Similarities and differences between training and testing tasks were noted above.

Significant statistical differences between treatment groups were not found by Treffinger & Ripple on transfer tasks. While claiming the tests were valid ones, Treffinger & Ripple acknowledged that the tasks were severe and rigorous measures of transfer. Performance on an *Arithmetic Problem Solving Test* and an *Arithmetic Puzzles Test*, for example, seemed very remote from the experiences provided by the *Productive Thinking Program*, the basis of the training. Success with such remote transfer tasks would give course designers considerable confidence about the adequacy of their program yet, unless claims were being made about the relevance of the program across subject areas, such rigor was unnecessary. Indeed, such harsh measures could be insensitive to some facets of transfer that did occur. This would suggest that a variety of measures ranging across the transfer dimensions nominated by Robinson *et al.* (1972) and Snow (1974) need to be applied in any training study if errors of interpretation are to be avoided.

While claiming that the Problem Tasks, *Indians* and *Famine*, were genuine transfer tasks, one reservation in interpreting the large level of statistical significance gained in the present research should be noted. Circumstances dicated that the two transfer tasks, plus the other two problems, were administered at the same testing session. A characteristic of strong measures of transfer, as discussed by Robinson *et al.* (1972), was that no cue should be given to the student to apply a method of problem-solving that had been taught. While in practice it may be difficult to hoodwink the student, it was true that there were strong cues present in the testing situation of the present investigation. Ideally, it would have been better to have separated the two sub-sets of Problem Tasks and had them administered by

different interviewers, thus giving a stronger indication of transfer. Or perhaps (among yet other alternatives) to have the regular classroom teacher give the transfer tests.

Was the Training Success Durable?

At a re-testing on Problem Tasks some eight weeks after the post-testing session, the observed differences between treatment group means were not statistically significant. The treatment effects originally identified at the post-testing session had dissipated. When presenting this result in the previous chapter, it was noted that the multivariate statistic for treatment effect was significant. However, this result was confounded by a significant interaction between treatment and school. In addition, the univariate analysis indicated that treatment effects were only significant on one out of four measures. Therefore, in general, the gains in performance arising from the special training experiences were not seen as lasting.

The interaction effects between treatment and school, suggesting durability of training in some schools, was confusing. It was possible that the activities offered in some schools, after the completion of the experimental program, complemented what had been previously done in the training sessions. Indeed, some class teachers expressed interest in the training programs and were shown lesson plans at the completion of the post-testing program. However, it was more likely that such interest would have been reflected in a class by treatment interaction rather than a school by treatment one. Furthermore, there appeared to be no satisfactory explanation for durability of performance with one treatment group rather than the other, particularly as it was the weaker treatment, Treatment Two, which had the better survival record. It seemed unlikely that the experiences provided within a school would complement one treatment group without at the same time reinforcing the other. Was it possible that differences in the attitudes of subjects towards the enquiry process were engendered by the different training programs? Did Treatment Two, the more open discussion situation, develop a better attitude within the children towards the enquiry process and thus aid the durability of the experiences? The evidence was too flimsy to even hazard a guess. Further research is necessary.

While the lack of durability of treatment effects was disappointing, it was perhaps unrealistic to expect anything else. The training provided was narrow in scope, being restricted to problems set in the context of Social Studies photographs. Moreover, the duration of the experience was rela-

tively short. Although much longer than the hour of instruction time which Snow (1974) asserted as the time resource most frequently allocated in experiments, it was short considering the complexity and multiplicity of behaviours the program was expected to influence. With more time, with a program using a greater diversity of materials and variety of activities, and with the underlying operative schemes being employed in other areas of the curriculum, it seemed reasonable to expect that gains noted on the post-testing sessions would become durable. Indeed, in such circumstances it would not be unexpected to find even greater improvement in performance.

Few of the studies training children in enquiry concerned themselves with the durability of the identified treatment effects. Of those that did, a number reported treatment effects persisting weeks after the completion of training (Elsmere, 1963; Olton & Crutchfield, 1969; Scott, 1973). Olton & Crutchfield reported the presence of training effects more than six months after training had finished. Even more incredible was Scott's finding that some aspects of the initial success achieved by a training program persisted some six years later. These results were in marked contrast to those of the present investigation. Yet on reflection, it will be recalled that the program provided in the Olton & Crutchfield study had more variety in its activities, and the time devoted to its consideration was more than double that given in the present study. The Scott program used the enquiry strategy conceived by Suchman (1961). Students were exposed to this program over two or three years. With both Blatt & Kohlberg (1973) and Elsmere (1963) 12 weeks of training was provided. Where shorter periods of training were given, at the time of retesting, the training effects originally noted had dissipated (Brederman, 1973).

Why was there a Lack of Improvement on some of the Component Measures of the Enquiry Process?

Special attention was given in the training program of Treatment One to the development of a number of skills underlying successful problem-solving. These mental schemes included: making inferences from events portrayed in photographs, creating numerous solutions to given problems and searching photographs for evidence to justify a given answer. Despite the special attention, the performance of the treatment group was not seen as differing significantly on measures evaluating these abilities from that group of subjects who received no training. Did the program fail? Were there alternative explanations for the lack of significant differences between the training group and the control on component measures of enquiry?

Rather than seeing weakness in a training program, Wardrop *et al.* (1969), in their study, linked the failure of some tests to differentiate between treatment and control groups to the instruments themselves. The offending tests were structured to elicit a large number of rather short, discrete responses on relatively uncomplicated tasks. Wardrop *et al.* claimed that these tasks did not reflect the intent of the program. What was learned in training was only marginally relevant to performance on these brief problems. Attributing the absence of observable performance differences between trained and non-trained groups of subjects to deficiencies in test instruments did not seem appropriate in the present study. True, some of the tests used assessed the child's ability to produce a large number of a particular type of response. True, the tests aimed to examine the child's achievement on specific elements of the complex enquiry process. But it was the specific intention of the program to enhance these particular abilities. The component tests were included to help identify the factors contributing to treatment effects observed on the complex Problem Tasks.

The component tests had shown in trials that they were capable of registering performance gains. Adults to whom the tests were administered in the developmental phase scored considerably better than grade six students. Unlike most individuals in the Jenkinson & Lampard (1959) study, the subjects of the present investigation were not at the ceiling of the test scale prior to the commencement of training. Furthermore, the trials showed that the component test had moderately high levels of internal consistency and could be scored reliably. This information was reported in Chapter 6. It therefore seemed inappropriate to attribute the failure of the component tests to discriminate between training and non-training groups to deficiencies in the test instruments.

An alternative explanation for the lack of treatment impact centred on the performance levels of children prior to training. Perhaps most children at grade six level already had mastery of the component skills of problem-solving being measured by the Problem Tests. Instruction, therefore, would have been irrelevant as training could not have significantly changed performance.

A number of studies have identified that many children in the upper levels of the primary school do have many of the component skills associated with problem-solving (Allender, 1969; Wardrop *et al.*, 1969; Jurd, 1973; Wicks, 1974). For example, Wicks' research indicated that seven-year-old children were able to generate alternative solutions to Social Studies problems. Earlier, Allender noted that children in Grades four, five and six were able to sense problems, identify the nature of each problem and search for informa-

tion to aid in decision-making. Yet other researchers have drawn attention to the inability of primary school children to employ these same skills. Lovell (1961), Duckworth (1972) and Gray (1972) all commented upon the difficulty children had in finding multiple answers to given problems. McNaughton (1966) reported the inability of 12-year-old children to see relationships within given data.

In part the divergent research findings could be attributed to the nature of the problems the child was being asked to solve and the level of mastery demanded. Lovell based his comments upon observations of subjects attempting to solve Piagetian problems such as flexibility of rods, oscillation of a pendulum and combinations of colourless chemical liquids. These problems were quite different from those used by Wicks where the children were asked to respond to a social situation. Vandalism, the issue used by Wicks, was within the personal experience of most children whereas the manipulation of chemicals or pendulums was not. Moreover, the different researchers accepted different levels of performance as being adequate. Lovell appeared to expect each subject to be able to identify and consider all logical combinations of four chemicals in solving the colourless liquids problem. On the other hand, Wicks accepted as adequate the fact that his subjects could suggest two alternative solutions from a much larger set of possibilities. With classification skills, Higgins (1974) reported that, while some children had developed certain abilities, it could not be assumed that eight-, nine- and ten-year-olds had fully mastered basic elements of classification.

In the present research, children included in the sample were able to generate alternative answers to problems, seek evidence to support possible answers, see some relationships in the information gathered and infer information from photographs but the performance levels were well short of future potential as judged by adult performance. By comparison with adult scores on the Element Tests, the achievement levels of the grade six children were only moderate. For example, students averaged three solutions to each of the 10 problems presented on the *Finding Answers Test*. With the *Puzzles Test* the average number of pieces of evidence was slightly higher than two. Consequently, the explanation that training made no impact on improving the subjects' component skills of problem-solving because mastery had been achieved prior to instruction was untenable.

Deficiencies in the training program were a more likely explanation for the failure of subjects in special training groups to out-perform their peers in the control. The experiences provided to develop selected operative schemes

were apparently either not sufficiently intensive or not distributed over a sufficient period of time, or both, to influence performance. No more than two consecutive sessions were used in Treatment One to specifically draw the attention of subjects to each of the operative schemes measured by the tests—*Answers*, *Puzzles*, *Higgins Two* and *Higgins Four*. It was true, however, that these skills were applied in combination in later training sessions where new problems were considered. But when contrasted with other research the overall time given to developing operative schemes was short (Wardrop *et al.* 1969; Scott, 1973; Higgins, 1974). While focusing only upon inference-drawing behaviour, the Higgins program, for example, required 16 sessions of 30 to 40 minutes duration for completion.

In the present study it has emerged that sufficient emphasis was not placed upon student productivity. Written responses collected by the teachers at the end of each training session indicated that only two or three suggested answers to any given problem were the norm. Evidence collected to support an answer only occasionally went beyond three pieces of information. Such limitations in responses were directly attributable to the program. When teachers used model responses to illustrate appropriate behaviour only two or three possible answers were detailed. Rarely were more than three pieces of information used in supporting particular solutions (see Notes of Lessons in Whitehead, 1975: Appendix A). Moreover, the duplicated paradigm used to record the results of each step of the enquiry provided space for only three possible answers and the area available for listing evidence was severely restricted. These factors inhibited improvement of the child's operative schemes, and hence, performance on the Element Tests.

In summary, the cause of the failure to improve a number of operative schemes associated with problem-solving appeared to lie with the shortness and lack of intensity of the training program. Explanations centring upon deficiencies in test instruments, or levels of student mastery prior to training, were not seen as being satisfactory.

Additional Comments on Treatments

Three additional points about the training programs need to be made because of their relevance to possible future development. First, all three experimental teachers agreed that the presentation of the overview of the enquiry process in the context of crime-solving was highly successful. Children in Treatment One were very much involved and in the teachers' view enjoyed

the activity of that particular session. It will be recalled that the crime-solving situation was used much more extensively with the *Productive Thinking Program* (Covington *et al.*, 1972).

Secondly, the use of a general enquiry paradigm as a record sheet was very successful. The paradigm was similar to that developed by Robinson *et al.* (1972). In the view of the teachers, it helped direct the attention of the students to the various facets of enquiry and aided their recall of appropriate detail. Significantly, however, no child included in the investigation used the pad and pencil provided in the test situation to record relevant information about the given problem. This was despite the emphasis in training sessions placed upon note-making as a memory aid and the use of the enquiry paradigm. The attention of all subjects was drawn to the pad and pencil in the test situation. Questions of interest are: 'If subjects had made notes would the quality of the responses improve?' and 'Why did they they not use the pad and pencil?'

Thirdly, with Treatment Two, improvement in student performance was expected to arise primarily from student interaction. Through debate and observation of the approaches adopted by other students each individual was expected to develop appropriate executive, figurative and operative schemes. In fact, teachers reported that, in discussion, there was very little effective interaction between students. Few students appeared willing to modify an opinion despite the presentation of convincing alternative solutions by other members in the group. The normal reaction to a conflicting position was to reiterate the earlier statement, often in a louder and more aggressive manner. Where a student, in presenting his case, was challenged by another student, the challenge appeared to be viewed as an irrelevant interruption and as such was ignored. Perhaps these students, being totally engrossed in their own argument, were utilizing all the M-space available to them and were therefore unable to build in additional ideas advanced by their classmates. A less tenuous explanation, perhaps, is the possibility that the Problem Tasks and the context in which they were presented were not sufficiently motivating for the subjects to learn.

9. Retrospect and Prospect

Conclusions: A Summary

The conclusions of the investigation can be summarized as follows:

1. The mean differences between both Treatment One and Two and the control were significant on the multivariate analysis of scores from the four Problem Tasks. The univariate analysis for each Problem Task also revealed significant differences between the two training groups and the control. Subsequent examination of the estimates of effects indicated a significant difference between the means of Treatment One and Treatment Two. The explicit instruction program was a much more powerful influence upon performance than the treatment which primarily relied upon interaction between students and the experience of doing the tasks. Training successes were attributed to improved executive schemes. As a result of the experiences provided in training sessions it was maintained that the subjects had a clearer view of the nature of the task. The possibility that improved operative or figurative schemes were major factors in the observed success was dismissed after examining the patterns of responses in the data.
2. Significant differences between the observed means of the three treatment groups were identified on the Concept Tests—*Poverty* and *City*. The performance of both Treatments One and Two were significantly better than Treatment Three. As with the Problem Tasks, Treatment One had a greater effect on performance than Treatment Two. Training success was attributed to the listing, grouping and labelling strategies, together with the opportunity given to apply the ideas gained to problems posed.

3. The null hypothesis expressing the absence of significant differences between the performance of the treatment groups on the three Element Tests—*Finding Answers, Puzzles* and *Working it Out*—was rejected. Large differences between treatment groups were seen to exist. From the univariate analysis it became clear that the major source of the significant multivariate statistic was performance on the *Working it Out Test*. On that test, the mean differences between Treatment One and the control was significant. The observed difference between Treatment Two and the control was not significant and thus accepted as a chance fluctuation. In the time available, the experience of engaging in the enquiry process without teacher direction was insufficient to improve the operative schemes associated with the *Working it Out Test*. Where explicit instruction was provided, significant gains were made.
4. The hypothesis that there would be no difference between the performances of the three treatment groups on the *Higgins Inference Tests* was sustained by the statistical analyses.
5. The observed differences between the performance of the various treatment groups on the ACER Social Studies Tests were accepted as being due to sample fluctuations. The null hypothesis was accepted. The training programs were contributing nothing of significance to performance on Social Studies terminology or comprehension of Social Studies material as measured by the ACER tests.
6. The null hypothesis associated with durability of training was accepted. Training effects were not seen to be durable on the four Problem Tasks. Although the relevant multivariate statistic was significant, the decision was made to ignore it. This position was taken because of a strong interaction effect with treatment effect. Moreover, the source of the significant multivariate statistic was performance on only one of four measures. Treatment effects were not accepted as being durable.
7. Statistically strong transfer effects were noted. The mean scores of the two special treatment groups differed significantly from that of the control. Treatment One subjects also performed significantly better on the two verbal Problem Tasks than the subjects in Treatment Two. Compared to Treatment Two the training effects of Treatment One were almost twice as great.
8. Significant teacher effects were noted in three analyses. These analyses concerned the Concept Tests, the Element Tests and the *Higgins Inference Tests*. Teacher effects on the ACER Social Studies Tests, although rejected as being not significant at the .05 level, were on the borderline

of acceptability. With the Concept Tests the teacher effect appeared to be located in performance on the *City* instrument. Statistical information failed to reveal the source of the significant teacher difference. With the Element Tests all three instruments had significant univariate teacher effects. On *Finding Answers* and *Puzzles* the groups associated with Teachers One and Two significantly out-performed those subjects with Teacher Three. Results from *Working it Out* showed that the mean score of the subjects working with Teacher One was significantly better than that of the students who worked with Teacher Three. However, with the *Working it Out Test* performance differences between the Teacher Two group and the Teacher Three group were accepted as chance rather than statistical differences. Strong univariate teacher effects were noted on both *Higgins Inference Tests*. Groups of subjects allocated to Teachers One and Two performed significantly better than those associated with Teacher Three.

There was not always a consistent pattern in teacher effects across the tests. Moreover, the intensity of the effects varied. For example, estimates of teacher effects indicated that Teacher One was twice as effective as Teacher Two on the *Puzzles* test but on the Higgins instruments the performance of the two groups were almost equal. Being outside the bounds of the investigation, the reason why such teacher effects should have occurred was not explored; although the possibility of initial group differences prior to training and differences in teacher instructional style were noted.

9. School effects were significant on three sets of measures. With *Finding Answers* and *Working it Out*, School Three performed significantly better than School Four. The same pattern was observed with the *Higgins Inference Tests* and the ACER Social Studies Tests. However, in addition, in the latter instance the performance of School One was significantly better than School Two. An examination of the characteristics of the schools concerned did not reveal the reason or reasons for the differences in performance. The only plausible explanation appeared to be that the students in one school compared to another did not put maximum effort into completing particular tests.
10. Classroom effects were significant on the multivariate analysis of the ACER Social Studies Tests. However, because neither univariate analysis proved significant, the source of the significant multivariate statistic was not explored. Other than at a superficial level, suitable information on class characteristics revealed no clue to the observed differences in performance.

11. Only one multivariate interaction proved significant. That was a Treatment by Teacher interaction on the three Element Tests. Such an interaction suggested that some teachers operated relatively better with one treatment in comparison to another. The source of the significant was the *Working it Out Test*. Teacher One performed significantly better with Treatment Two compared to Treatment Three, than did Teacher Three. Teacher Two was less effective with Treatment One compared to Treatment Three than was Teacher Three. This interaction between Treatment and Teacher was explained tentatively in terms of teacher style. Perhaps some teaching styles more comfortably accommodate certain teaching procedures.
12. The experience of doing the Problem Tasks on one occasion did not significantly influence performance on a subsequent re-testing. Two analyses were undertaken and both implied the same conclusion.
13. Although a significant relationship was noted between some of the component enquiry measures and the Problem Tasks, the relationships were described as moderate. A canonical correlation analysis indicated that of the four possible functions joining the two batteries of tests only the first was significant. The null hypothesis indicating no relationship between performance on the enquiry process tests and Problem Tasks, either individually or in linear combination, was rejected but it was noted that the relationship was not strong.

Limitations of the Study

Before considering the educational implications of the findings summarized above, it is necessary to recognize that two features of the present study limit the generalization of the results. These features were the attributes of the enquiry process being measured by the test instruments and the sample of students included in the investigation.

Various views of what skills constitute the enquiry process have been advanced by writers (Nay, 1971; Michaelis, 1973; Victoria, Education Department, 1974). All such facets were not explored in the present investigation. Some that were seen to be crucial were separated out for examination. Others were ignored in the testing program. Amongst the skills selected were the generation of possible answers, the gathering of evidence, inference-making behaviour, and the ability to weigh up evidence associated with alternative solutions. For example, attention was not given to measuring the ability of children to identify problems as it was in the studies of

Shulman (1965) and Allender (1969). With the Element Tests and the Problem Tasks the issues were already identified for the student. On the Problem Tests, to rate responses, attributes similar to those measured separately on the Element Test were used in combination. No attempt was made to assess the quality or correctness of the solutions provided, as for example, in the studies of Wardrop *et al.* (1969) or Treffinger & Ripple (1970). It should, however, be pointed out that few investigations have measured performance against multiple criteria of any description.

The sample was drawn from sixth grade students in six primary schools. The schools were located in middle-class areas of metropolitan Melbourne. Inner-suburban schools, country schools and schools in lower socio-economic areas were not considered. Free of serious language problems, the schools selected were staffed by teachers and principals who gave their full co-operation to the project. Such schools did not provide a satisfactory base from which to generalize the findings of the present study to primary schools throughout Victoria. Mason (1963) for example, found that training success was influenced by the grade level of the subjects. As grade level increased, Ripple & Dacey (1967) found instructional materials became less effective in developing creative thinking abilities. Brydon (1967), cited in Peel (1971), and Poole (1973) reported that cognitive performance was influenced by the socio-economic level of the subjects. Both Taba (1964) and Whitehead (1972) found a very low relationship between performance on Social Studies Problem Tasks and socio-economic level but this could have been due to the lack of precision in measures of socio-economic status. Because of the constraints of the sample, a fixed effects model was employed in the statistical analyses. However, it could be argued that similar results to those obtained in this research might be expected in schools which exhibited similar characteristics to those used in the investigation (Bracht & Glass, 1968).

Suggestions for Further Research

While providing information on some questions, research studies tend to generate additional questions. Some of these arise from the results of the investigation and others from what are judged to be limitations of the original project. The present study was no exception. Some suggestions for further research are therefore listed.

1. To examine the generality of the results of the present investigation it would be useful to replicate the study. In such replication, factors such as socio-economic status and sex of subjects, school location, classroom environment and class level should be taken into account. Brydon (1967) and Poole (1973) identified a significant link between socio-economic class and cognitive style. Would the training provided in the present investigation be equally successful with subjects drawn from lower socio-economic families? Wardrop *et al.* (1969) and Johnston (1972) reported significant differences in performance levels between boys and girls. Girls out-performed boys on Piagetian tasks in the Dale (1970) investigation. Do girls respond better to training than boys? It is possible that rural children do not perform as well on some cognitive tasks as urban children. Would rural, inner city and suburban respond equally well to enquiry training?

   The environments of 12 classrooms used in the present investigation tended to be similar. In the main, they were towards the traditional end of the *Dimensions of Schools* scale (Traub *et al.*, 1972). Would children from more open educational environments have benefited from the programs to the same degree as the subjects in the present investigation? In an educational climate where administrators are espousing what is labelled 'open education' and building schools on 'open plans' the question becomes one of some importance.

   Is it possible that children nurtured in the open educational style will respond better to the freer strategy adopted in Treatment Two of this project, than to the teacher directed approach of Treatment One? The Traub instrument could be used to help identify appropriate schools but it would be necessary to make some modifications to bring the questionnaire more into line with local practices. Some of the items would fail to discriminate between schools.

2. Investigations could also be directed towards exploring the relationship between performance on Social Studies Problem Tasks and component measures on the enquiry process. In the present investigation the relationship was only moderate. Why should this be? Were there some enquiry abilities not being surveyed in the current battery of tests? After an analysis of the abilities being measured on the tests employed in this research, new instruments could be developed for use in future training studies. By this means, the belief that performance on the Problem Tasks entails more than the summing of achievement on component skills could be examined more closely.

3. The highly significant differences between treatment groups gained in the present study were attributed to improved executive schemes rather than to operative schemes. With the research of Olton&Crutchfield (1969) and Scott (1973) successes were associated with alerted operative schemes. Both these latter studies achieved some measure of durability. Was it conceivable, given only a short training period, that executive schemes were more susceptible to fading than operative schemes? Perhaps this was a factor explaining the presence of durability in some studies and its absence in others. In further research attention should be given to the amount of re-training necessary to re-alert appropriate executive schemes. In terms of training time, how much effort is required to reactivate dormant executive schemes?
4. The question of transfer was one requiring further research. Tests demanding more distant generalization of enquiry skills than those used in the present investigation are necessary. Such research would add to knowledge concerning the existence of specific and/or general modes of thought: an issue not yet adequately resolved (Stone & Ausubel, 1969; Johnston, 1972; Lovell, 1971). Tests similar to those used by Case & Fry (1974) in the physical sciences could be used together with the more rigorous transfer tasks of Ripple & Dacey (1967) and Treffinger & Ripple (1970).
5. A question of importance to the Social Studies curriculum designer is the extent to which the subjects trained in enquiry used that procedure outside the classroom. Robinson *et al.* (1972) noted the significance of this question. An assumption behind the recommendation of enquiry objectives in Social Studies is that students will adopt rational procedures when examining real problems of concern to themselves and their community. Do students employ the problem-solving strategies as part of their everyday method of operation? In the present research the explicit instruction proved to be a considerably more powerful influence upon performance than the approach where children discussed problems without teacher intervention. Is there a significant difference between these two programs in the attitudes they engender in the students towards applying enquiry strategies? To help answer these questions a radio transmitter could be used to 'bug' the discussion of children outside school hours. Cambourne (1973) used this technique to study the verbal behaviour of grade one children. Obviously, such an investigation would be fraught with dangers and difficulties but with child agreement, parental co-operation and a little structuring of situations, not unrealistic.

6. The theoretical base for the training programs used in the present research was that developed by Pascual-Leone (1970). Two tests—*Backward Digits* and the *Group Embedded Figures Test*—used in the present study related to that theory. The first measure pertained to the notion of M-space and the second to the construct of field-dependence. Performance on these tests did not significantly relate to performance on the Problem Tasks or component tests of enquiry. Why should this be? Saarni (1973), after noting performance on the rod and frame tests from the *Productive Thinking Program* (Covington *et al.*, 1972), reported the view that field-dependence had doubtful implications for complex problem-solving performance. Case (1972), using the WISC blocks problem as a measure of field-dependence, interpreted the lack of a high correlation between that test and another task as the lack of misleading clues in the latter instrument. This explanation fails to account for the lack of significant relationships between measures in the present study because the Social Studies tasks were not devoid of misleading cues. Research is needed to establish the nature of the relationship between performance on Social Studies problem-solving tasks and the theoretical constructs of M-space and field-dependence.
7. The present research focused upon an assumption underlying modern Social Studies programs, namely that the rational operation of students can benefit from training. There are other important assumptions that this research has not examined. These refer to the relationship between improved enquiry skills and better decision-making; the relationship between rational enquiry and independence of thought and action; and the relationship between logical and systematic enquiry and the development of student empathy with persons of groups and cultures differing from their own. Research is required into whether these relationships actually exist.

Some Educational Implications

The findings summarized above lend strong support to the view that Social Studies problem-solving behaviour of grade six children can be improved through training. Such improvements are gained by presenting the student with a model of the relationships existing between various facets of enquiry and giving him the opportunity to apply the various skills in solving relatively simple problems. Current Social Studies programs sometimes provide, for the benefit of the teacher, an outline of the possible steps in a social

enquiry. No such provision is made for the child. The child is led through the steps of enquiry by skilful teacher questioning. At no time is the child given an overview of the procedure (Taba *et al.*, 1971; Queensland, Education Department, 1972; Victoria, Education Department, 1974). This practice contrasts with that of the enquiry program developed by Robinson *et al.* (1972), where the starting point of the activities was an introduction to a general enquiry paradigm. The findings of the present research support such a procedure. Presentation of the paradigm in the context of crime-solving strategies proved popular with students and teachers in the present investigation so could well be repeated on future occasions. The crime-solving setting was used successfully, although more extensively, with grades five and six in the *Productive Thinking Program* (Covington *et al.*, 1972).

The relative success of Treatment One compared to Treatment Two suggested that carefully structured experiences are more beneficial to the enhancement of problem-solving performance than situations where the child is left to himself to discover appropriate actions. More attention, therefore, should be given to class instruction in the teaching of enquiry. This view is contrary to the intentions of some programs currently used in Victorian primary schools, The *Primary Science Curriculum Guide* (Victoria, Education Department, 1969), for example, pointed out that a teaching method emphasizing instruction was not desirable. The controversy related to discovery learning, guided discovery, and expository teaching has raged for some years (e.g. Friedlander, 1965; Shulman & Keislar, 1966; Herman, 1969). In the short term, however, greater gains in the problem-solving behaviour of grade six students appear more likely where direct instruction is given in the appropriate skills.

There is a place in Social Studies programs for the examination of relatively simple questions. Similar questions to those used in the training-sessions of the present research would be appropriate at the senior primary and junior secondary levels. Admittedly, this recommendation limits the time available to pursue more significant social problems. On the other hand, however, simple questions provide a better opportunity to view and apply the enquiry process as a totality. Children, when submerged in complex issues, often narrow their focus to small facets of the enquiry and fail to see the relationships between what they are currently doing, what has been completed and what is still to come.

Previous research indicated that some of the skills or operative schemes associated with enquiry were susceptible to training (Olton & Crutchfield, 1969; Wallen *et al.*, 1969; Wardrop *et al.*, 1969; Higgins, 1974). Such skills

included the ability to infer information from data, the production of ideas, and the ability to sense discrepant facts. The present research supported the general conclusion of the earlier studies. More specifically, the present investigation showed that the ability of students to manipulate data could be enhanced through training. This improvement arose from specific attention being given to the appropriate operative scheme. The teacher demonstrated to the students what was involved and opportunities were provided where they practised the skill. Where such attention was not provided and students were left to themselves to develop such a scheme there was no significant improvement. Social Studies curricula, in general, do not focus upon specific enquiry skills. In this regard, Social Studies programs differ from some mathematical programs where attention is directed to the diagnosis of weaknesses in particular computation skills, and to the provision of remedial assistance (Victoria, Education Department, 1968; Blackell, Goodger, Haig, Izard, Smith & Whitehead, 1969). Such provisions appear to be lacking in Social Studies programs. Where concern in Social Studies curriculum is with the development of enquiry abilities, it seems appropriate to provide skill-building programs; programs where attention can be focused upon specific skills; occasions where the child's limited cognitive processing capacity is not congested with a large number of competing mental schemes. Without such programs it seems unlikely, at least in the short term, to gain improvement in performance. Even where attention is directed to particular skills, care must be taken to identify the current level of performance and to ensure that the student is encouraged to surpass that degree of mastery. In the present research the activities associated with the generation of answers, the identification of evidence to support a particular solution, and the enhancement of inference-making behaviour were unsuccessful probably because sufficient attention was not given to the student's productivity.

As a result of the strategies employed in the training programs, the figurative schemes of the subjects were significantly improved. This was reflected in the performances on the Concept Tests. It will be recalled that both training programs in the present study encouraged the listing, grouping and labelling of information gathered from photographs. Although stress was not placed upon re-grouping and re-labelling, the strategy used was very similar to that advocated by Taba *et al.* (1971), used by the Queensland Education Department (1972), and developed further by the Victorian Education Department (1974). Because of this sameness the present findings provided some empirical support for the practices advocated in the concept

development phase of these Social Studies curricula. Furthermore, it would seem appropriate that the classification strategies should be adopted more widely in programs concerned with the enhancement of figurative schemes; programs such as Health Education and Science.

Since the treatment effects had dissipated some eight weeks after the completion of training, there was no justification for an uncritical acceptance of the programs used in the current research. In an effort to obtain durability and transfer of skills to problem areas employing different enquiry strategies, the program would need to be enriched. Photographs of social situations provided the focus for the problems examined in the present study. No attempt was made to build in direct experiences, movie film or book materials. Role play and simulation games were not used. All these data sources are exploited in modern Social Studies programs (Taba *et al.*, 1971; Whitehead, 1973; Victoria, Education Department, 1974). In addition, use could be made of programs which provide different types of problem situations (Covington *et al.*, 1972; Evans, Poole, Georgeff & Heffernan, 1974). Built into the normal school curriculum these activities would assist the child to see that problem-solving is not just concerned with Social Studies. Indeed, problem-solving should not be divorced from activities in Science, Mathematics, English, Health or Art. With an expanded thinking-skills program the activities could be spread more adequately throughout the school year, thus increasing the possibility of achieving durability.

Of the two programs used in the present research, one was consistently superior to the other on many measures. The better program was the one where the teacher had an initiating, instructing, demonstrating and integrating role. Other research has attributed stronger training effects to the active involvement of teachers (Covington, n.d.). At least in the short term, the present research suggested that better student performance was more likely where the teacher played a directive rather than supportive role. Should a teacher prefer to act more as a resource person in Social Studies, to be consulted by the children as required, then greater attention may have to be given to improving the communication skills of children. The general inability of subjects to listen to and to consider viewpoints that conflicted with their own was noted in the present research. Cognitive growth is unlikely to be achieved if the student fails to recognize and relate to the conflict existing between what he understands and what others understand. Again, an appropriate course of action would be for the teacher to assist the child to acquire the necessary communication skills. In this regard the work of Evans *et al.* (1974) warrants consideration.

Notwithstanding the highly significant statistical impact of the training programs, the actual level of performance achieved as a result of the experiences provided must be considered. In Piagetian terms, when it came to Social Studies-type problems, previous research described grade six students as concrete operators (Hallam, 1967; McNally, 1970; Johnston, 1972; Whitehead, 1972; Jurd, 1973; Nettle, 1975). Hallam and Jurd placed the commencement of concrete operations at 12 years, an average age for grade six children. Whitehead described grade six children as being in the early phase of concrete operations. As a result of training, the present research indicated that the average level of performance was still in the concrete stage of operations but at a more sophisticated level. Alternative answers to given problems were being considered and information was advanced to support them, but there were only very tentative efforts at the systematic examination of evidence to reach a final conclusion which is characteristic of the formal operational phase. Consequently the ability of these students to apply the full enquiry process as conceived by Dewey (1933) or more recently by Michaelis (1973) was still severely restricted despite the special training. Even where programs gave specific attention to the development of appropriate executive and operative schemes, the expectation that primary school and junior secondary school students will be able to apply the enquiry process as a totality seems extremely optimistic. As already indicated, Social Studies programs, in the main, do not provide such intensive experiences. Therefore, to proceed further and introduce primary or even junior secondary students to the subtleties of the process and the differences that exist in its application between different areas of knowledge does not seem warranted (Berlak, 1965; Connelly, 1969; Herron, 1971; Blachford, 1973). For primary school children, it is the similarities in the process across subject areas that need emphasis. Let the differences be highlighted later in the secondary school experiences of the child as is done, for example, in the program of Robinson *et al.* (1972).

The final educational implication relates to the practice of State Education Departments of surveying educational achievement in various subject areas. Tests have been specially developed for this purpose. Two tests designed and used by the Victorian Education Department to assess achievement in Social Studies were used in the present investigation—*Words in Social Studies* and *Comprehension in Social Studies* (Renehan & Wilkes, 1973). Performance on both instruments correlated at a low level with both the Problem Tasks and the Element Tests. Assuming the Problem Tasks and component tests reflect objectives of teaching Social Studies, it appears likely that the current survey program used in Victoria is too restricted. The

ACER Social Studies Tests should therefore be supplemented by other instruments. Additions could include tests like the Problem Tasks and component measures tests used in the present research.

In summary, the ability of grade six children to employ the enquiry process can be improved. Such improvement can be achieved by drawing the child's attention to an overview of the process. Both figurative and operative schemes can benefit from class experiences. In the short term at least, improved executive, figurative and operative schemes are more likely to emerge where the teacher is actively engaged in instructing and demonstrating appropriate behaviour than a program where students are left to identify appropriate actions for themselves. In addition, where it is desirable for problem-solving skills to be lasting and transferable to new tasks it appears likely that training will need to be applied consistently throughout the school year and across different subject areas. Considering the sophistication of grade six students in applying the enquiry process, attention might best be directed towards the similarities in the process across subject areas rather than the differences.

Concluding Remarks

Recently Social Studies programs have come under close scrutiny in Australia. New curricula have been and are being developed in all States at both the primary and secondary school levels. The current research has brought forward evidence of importance to teachers and curriculum workers engaged in these developments. From the results of the study, empirical support has been gained for some of the teaching practices advocated; a cautionary note sounded about the over-optimistic objectives of some programs; and indications given of act[redacted] that could profitably be included in subsequent curriculum developments.

The impact of two training programs upon the cognitive performance of grade six children has been examined systematically. Answers were gained to the following questions. Can students be trained to apply the enquiry process? What are the components of the training which account for success? Did the training effects transfer to tasks different from those experienced in class? Did the training effects persist some eight weeks after the completion of training and what was the relationship between performance on problem tasks and performance on component skills of enquiry? Hopefully the answers to and discussion of these questions will contribute to improved classroom practices in the teaching of Social Studies.

Appendix

TABLE 9
Sequence of Effects in the Analysis of Variance

Source
Constant
Treatment, eliminating constant
Teacher, eliminating constant and treatment
School within teacher, eliminating constant, treatment, and teacher
Class nested within school and teacher, eliminating constant and all main effects
Treatment x teacher, eliminating all above
Treatment x school/teacher, eliminating all above
Treatment x class/school/teacher, eliminating all else

TABLE 10
Description of Data on all Criterion Measures in Terms of Treatment Groups

	Poverty			City			Indians			Famine			ACER Words			ACER Comprehension		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
N	70	71	71	70	71	71	70	71	71	70	71	71	69	68	69	69	68	69
$\bar{X}$	4.03	3.44	2.70	4.14	3.23	2.44	3.44	2.58	1.75	3.66	2.62	2.09	25.32	25.06	24.32	25.22	24.57	25.01
SD	1.85	1.72	1.36	1.95	1.92	1.93	2.21	2.10	1.81	1.97	1.85	1.69	6.32	6.69	6.91	6.15	6.75	9.32
Kurtosis	-1.56	-1.24	1.21	-0.36	-1.15	-1.36	-1.44	-0.81	-0.32	-1.35	-0.10	0.73	0.14	-0.56	0.41	-0.37	0.03	-0.54
Skewness	-0.14	0.48	1.66	-0.79	-0.34	0.29	-0.05	0.61	0.91	0.18	0.84	0.98	-0.66	-0.52	-0.65	0.46	0.57	0.34

	Answers			Puzzles			Working It Out			Concept Poverty			Concept City			Higgins 2			Higgins 4		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
N	63	63	57	63	63	57	63	63	57	70	71	69	70	71	69	67	67	66	67	67	66
$\bar{X}$	31.50	28.54	31.33	25.33	21.29	24.21	22.79	17.02	16.86	8.60	7.94	7.44	9.47	8.89	8.19	10.28	9.87	10.54	9.96	9.78	10.35
SD	8.33	9.29	7.64	8.38	8.30	8.70	5.77	6.01	6.09	1.79	1.55	1.68	2.21	1.83	1.97	4.11	4.57	4.24	4.47	4.34	4.32
Kurtosis	3.12	1.40	-0.04	-0.44	2.18	-0.42	-0.16	0.27	1.19	-0.23	0.27	-0.22	0.21	-0.26	0.86	-0.75	-1.06	-0.98	0.66	-1.03	0.65
Skewness	-1.38	-0.57	0.21	0.01	0.84	0.50	0.58	0.89	0.89	0.07	-0.32	0.07	-0.58	-0.02	0.01	-0.23	0.27	0.05	0.60	0.05	0.31

TABLE 11

Summary of Analysis of Variance for Four Problem Tasks

		Simultaneous Test			Univariate F Statistic			
Source	df	Λ	F	df	*Poverty*	*City*	*Indians*	*Famine*
Constant	1							
Treatment	2	0.72	7.80***	(8,344)	11.19***	13.14***	11.92***	14.16***
Teacher	2	0.94	1.26	(8,344)	0.14	0.40	4.46*	1.86
School w teacher	3	0.93	1.13	(12,455.36)	0.59	1.92	0.65	1.65
Class w school/teacher	6	0.87	0.98	(24,601.2)	1.58	0.94	0.27	1.55
Treatment x teacher	4	0.96	0.45	(16,526.1)	0.20	0.23	1.28	0.23
Treatment x school	6	0.85	1.22	(24,601.2)	1.24	1.09	0.65	1.70
Treatment x class	12	0.72	1.24	(48,664.6)	1.35	0.92	1.34	2.14*
					Mean Squares			
Within	175				2.71	3.79	4.12	3.08
Total	N=211							

***significant at p<.001
** significant at p<.01
* significant at p<.05

TABLE 12

Estimates of Effects Between Treatment Groups on Problem Task: *Poverty*

	Estimated Means	Estimates of Effects	Standard Error of Estimation	Estimates of Effects in S.D. units
Treatment 1	4.03	1.31*	0.28	0.80
Treatment 2	3.44	0.72*	0.28	0.44
Treatment 3	2.71			

*significant at $p<.05$

TABLE 13

Estimates of Effects Between Treatment Groups on Problem Task: *City*

	Estimated Means	Estimates of Effects	Standard Error of Estimation	Estimates of Effects in S.D. units
Treatment 1	4.14*	1.69*	0.33	0.87
Treatment 2	3.23	0.77*	0.33	0.39
Treatment 3	2.46			

*significant at $p<.05$

TABLE 14

Estimates of Effects Between Treatment Groups on Problem Task: *Indians*

	Estimated Means	Estimates of Effects	Standard Error of Estimation	Estimates of Effects in S.D. units
Treatment 1	3.44	1.69*	0.30	0.83
Treatment 2	2.58	0.82*	0.34	0.40
Treatment 3	1.76			

*significant at $p<.05$

TABLE 15

Estimates of Effects Between Treatment Groups on Problem Task: *Famine*

	Estimated Means	Estimates of Effects	Standard Error of Estimation	Estimates of Effects in S.D. units
Treatment 1	3.66	1.57*	0.30	0.90
Treatment 2	2.62	0.53*	0.30	0.30
Treatment 3	2.09			

*significant at $p<.05$

TABLE 16

Summary of Analysis of Variance for Two Concept Tests: *Poverty* and *City*

		Simultaneous Test			Univariate F Statistic	
Source	df	Λ	F	df	*Poverty*	*City*
Constant	1					
Treatment, eliminating constant, teacher, school and class	2	0.87	6.39***	(4,344)	8.45***	7.37***
Teacher	2	0.95	2.40*	(4,344)	0.46	3.75*
School w teacher	3	0.96	1.33	(6,344)	0.59	2.15
Class w school/ teacher	6	0.95	0.79	(12,344)	0.22	1.27
Treatment x teacher	4	0.94	1.47	(8,344)	2.14	0.98
Treatment x school	6	0.94	0.97	(12,344)	1.13	0.86
Treatment x class	12	0.86	1.09	(24,344)	1.52	0.68
					Mean Squares	
Within	173				2.75	3.93
Total	N=209					

***significant at $p<.001$
** significant at $p<.01$
* significant at $p<.05$

TABLE 17

Estimates of Effects Between Treatment Groups on The Concept Test: *Poverty*

	Estimated Means	Estimates of Effects	Standard Error of Estimation	Estimates of Effects in S.D. units
Treatment 1	8.60	1.16*	0.28	0.70
Treatment 2	7.95	0.51*	0.28	0.31
Treatment 3	7.44			

*significant at $p<.05$

TABLE 18

Estimates of Effects Between Treatment Groups on The Concept Test: *City*

	Estimated Means	Estimates of Effects	Standard Error of Estimation	Estimates of Effects in S.D. units
Treatment 1	9.47	1.29*	0.34	0.65
Treatment 2	8.88	0.70*	0.34	0.35
Treatment 3	8.18			

*significant at $p<.05$

TABLE 19

Estimates of Effects Between Teachers on The Concept Test: *City*

	Estimated Means	Estimates of Effects	Standard Error of Estimation	Estimation of Effects in S.D. units
Teacher 1	9.29	0.41	0.34	0.21
Teacher 2	8.37	–0.51	0.33	–0.26
Teacher 3	8.88			

TABLE 20

Summary of Analysis of Variance for Element Tests: *Answers, Puzzles* and *Working It Out*

Source	df	Simultaneous Test			Univariate F Statistic		
		Λ	F	df	*Answers*	*Puzzles*	*Working It Out*
Constant	1						
Treatment eliminating constant, teacher school and class	2	0.69	9.71***	(6,288)	2.56	4.49*	27.64***
Teacher- eliminating constant, treatment, school and class	2	0.65	11.34***	(6,288)	13.86***	21.26***	25.26***
School w teacher	3	0.75	4.87***	(9,350.6)	4.96**	2.26	3.01*
Class w school /teacher	6	0.83	1.50	(18,407.8)	0.55	3.12**	2.39*
Treatment x teacher	4	0.85	1.97*	(12,381.3)	2.08	0.28	4.34**
Treatment x school	6	0.90	0.83	(18,407.8)	1.44	1.08	0.37
Treatment x class	12	0.85	0.67	(36,426.2)	0.59	0.37	0.81
					Mean Squares		
Within	146				50.35	47.12	25.92
Total	N=182						

***significant at p<.001
** significant at p<.01
* significant at p<.05

TABLE 21

Estimates of Effects Between Treatments on Element Test: *Working It Out*

	Estimated Means	Estimates of Effects	Standard Error of Estimation	Estimation of Effects in S.D. units
Treatment 1	22.98	6.29*	0.93	1.24
Treatment 2	17.15	0.47	0.93	0.09
Treatment 3	16.69			

*significant of $p<.05$

TABLE 22

Estimates of Effects Between Treatments On Element Test: *Puzzles*

	Estimated Means	Estimates of Effects	Standard Error of Estimation	Estimation of Effects in S.D. units
Treatment 1	25.36	1.41	1.25	0.20
Treatment 2	21.53	−2.42*	1.26	−0.35
Treatment 3	23.95			

*significant at $p<.05$

TABLE 23

Estimates of Effects Between Teachers on Element Test: *Answers*

	Estimated Means	Estimates of Effects	Standard Error of Estimation	Estimation of Effects in S.D. units
Teacher 1	33.81	6.50*	1.29	0.92
Teacher 2	31.68	4.37*	1.29	0.62
Teacher 3	27.31			

*significant at $p<.05$

TABLE 24

Estimates of Effects Between Teachers on Element Test: *Puzzles*

	Estimated Means	Estimates of Effects	Standard Error of Estimation	Estimation of Effects in S.D. units
Teacher 1	28.05	8.13*	1.25	1.18
Teacher 2	22.88	2.97*	1.25	0.43
Teacher 3	19.92			

*significant at $p<.05$

TABLE 25

Estimates of Effects Between Teachers on Element Test: *Working It Out*

	Estimated Means	Estimates of Effects	Standard Error of Estimation	Estimation of Effects in S.D. units
Teacher 1	22.68	6.22*	0.92	1.22
Teacher 2	17.66	1.19	0.93	0.23
Teacher 3	16.48			

*significant at $p<.05$

TABLE 26

Estimates of Effects Between Schools Nested Within Teachers on *Answers Test*

	Estimated Means	Estimates of Effects	Standard Error of Estimation	Estimation of Effects in S.D. units
School 1	34.94	2.26	1.85	0.32
School 2	32.68			
School 3	34.95	6.55*	1.86	0.92
School 4	28.41			
School 5	26.53	−1.56	1.79	−0.22
School 6	28.09			

*significant at $p<.05$

TABLE 27

Estimates of Effects Between Schools Nested Within Teachers on *Working It Out*

	Estimated Means	Estimates of Effects	Standard Error of Estimation	Estimation of Effects in S.D. units
School 1	23.79	2.21	1.33	0.43
School 2	21.58			
School 3	19.17	3.02*	1.33	0.59
School 4	16.15			
School 5	16.95	0.93	1.28	0.18
School 6	16.01			

*significant at $p < .05$

TABLE 28

Estimated Means for Teacher X Treatment Interaction on *Working It Out*

	Teacher 1	Teacher 2	Teacher 3
Treatment 1	27.61	19.45	21.89
Treatment 2	22.01	16.72	12.73
Treatment 3	18.45	16.82	14.81

TABLE 29

Estimates of Effects for Teacher X Treatment Interaction on *Working It Out*

	Teacher 1	Teacher 2	Teacher 3
Treatment 1	2.06*	-4.47*	
Treatment 2	5.61*	1.96	
Treatment 3			

*significant at $p < .05$

TABLE 30

Standard Errors of Estimation for Teacher X Treatment Interaction on *Working It Out*

	Teacher 1	Teacher 2	Teacher 3
Treatment 1	2.27	2.23	
Treatment 2	2.28	2.27	
Treatment 3			

TABLE 31

Summary of Analysis of Variance for *Higgins Inference Tests*

Source	df	Simultaneous Test			Univariate F Statistic	
			F	df	*Higgins 2*	*Higgins 4*
Constant	1					
Treatment, eliminating constant, teacher, school and class	2	0.99	0.39	(4,326)	0.63	0.46
Teacher, eliminating constant, treatment, and school	2	0.78	11.05***	(4,326)	19.12***	15.22***
School w teacher	3	0.82	5.54***	(6,326)	2.69*	9.18***
Class w school/teacher	6	0.92	1.09	(12,326)	0.41	1.20
Treatment x teacher	4	0.96	0.94	(8,326)	0.89	1.51
Treatment x school	6	0.95	0.76	(12,326)	1.31	0.58
Treatment x class	12	0.88	0.93	(24,326)	1.00	0.59
					Mean Squares	
Within	164				15.43	15.30
Total	N = 200					

***significant at p < .001
** significant at p < .01
* significant at p < .05

TABLE 32

Estimates of Effects Between Teachers on *Higgins Inference Test Two*

	Estimated Means	Estimates of Effects	Standard Error of Estimation	Estimates of Effects in S.D. units
Teacher 1	11.43	3.60*	0.71	0.89
Teacher 2	11.32	3.49*	0.70	0.86
Teacher 3	7.83			

*significant at p < .05

TABLE 33

Estimates of Effects Between Teachers on *Higgins Inference Test Four*

	Estimated Means	Estimates of Effects	Standard Error of Estimation	Estimates of Effects in S.D. units
Teacher 1	11.47	3.44*	0.70	0.85
Teacher 2	10.60	2.57*	0.70	0.64
Teacher 3	8.03			

*significant at p < .05

TABLE 34

Estimates of Effects Between Schools on *Higgins Inference Test Two*

	Estimated Means	Estimates of Effects	Standard Error of Estimation	Estimates of Effects in S.D. units
School 1	10.97	–0.91	1.00	–0.22
School 2	11.88			
School 3	12.21	1.78	0.99	0.44
School 4	10.43			
School 5	8.66	1.66	0.99	0.41
School 6	7.00			

TABLE 35

Estimates of Effects Between Schools on *Higgins Inference Test Four*

	Estimated Means	Estimates of Effects	Standard Error of Estimation	Estimates of Effects in S.D. units
School 1	10.97	–1.00	1.00	–0.25
School 2	11.97			
School 3	13.00	4.80*	0.98	1.19
School 4	8.20			
School 5	7.77	–0.51	0.99	–0.13
School 6	8.28			

*significant at $p<.05$

TABLE 36

Summary of Analysis of Variance for ACER Social Studies Tests

Source	df	Simultaneous Test			Univariate F Statistic	
		Λ	F	df	*Words*	*Comprehension*
Constant	1					
Treatment, eliminating constant, teacher, school and class	2	0.99	0.44	(4,336)	0.29	0.08
Teacher, eliminating constant, treatment, school and class	2	0.95	2.39	(4,336)	3.85*	3.53*
School w teacher eliminating constant, treatment, teacher and class	3	0.91	2.78*	(6,336)	1.09	4.03**
Class w school/teacher	6	0.88	1.93*	(12,336)	0.71	0.82
Treatment x teacher	4	0.96	0.82	(8,336)	0.94	1.44
Treatment x school	6	0.95	0.66	(12,336)	0.35	0.50
Treatment x class	12	0.89	0.87	(24,336)	0.61	0.29
					Mean Squares	
Within	169				44.95	75.03
Total	N=205					

***significant at p<.001
** significant at p<.01
* significant at p<.05

TABLE 37

Estimates of Effects Between Schools on ACER Social Studies Test: *Comprehension*

	Estimated Means	Estimates of Effects	Standard Error of Estimation	Estimates of Effects in S.D. units
School 1	28.99	4.67*	2.14	0.54
School 2	24.32			
School 3	28.21	4.82*	2.09	0.56
School 4	23.39			
School 5	24.42	3.03	2.07	0.35
School 6	21.40			

*significant at p<.05

TABLE 38

Summary of Analysis of Variance on Problem Tasks at Delayed Post-Testing

		Simultaneous Test			Univariate F Statistic			
Source	df	Λ	F	df	*Poverty*	*Slums*	*Indians*	*Gold*
Constant	1							
Treatment, eliminating constant, teacher, school and class	2	0.77	2.83**	(8,158)	10.29***	3.08	1.18	1.69
Teacher	2	0.71	3.72***	(8,158)	0.20	2.01	1.61	10.83***
School w teacher	3	0.84	1.17	(12,209.3)	1.03	2.35	0.54	2.06
Treatment x teacher	4	0.82	1.03	(16,241.9)	1.55	0.43	1.47	2.13
Treatment x school	6	0.56	2.08**	(24,276.8)	2.77*	1.28	1.25	3.33**
						Mean Squares		
Within	82				2.30	2.76	5.15	2.57
Total	N = 100							

***significant at p = .001
** significant at p = .01
* significant at p = .05

TABLE 39

Estimated Means on Problem Task: *Poverty* at Delayed Administration

	Schools					
	1	2	3	4	5	6
Treatment 1	4.40	4.69	4.06	4.63	4.38	4.71
Treatment 2	3.01	4.43	4.61	2.44	4.06	3.39
Treatment 3	3.74	2.04	1.90	3.50	2.73	3.06

TABLE 40

Estimates of Effects and Standard Error of Estimation Between Treatments x Schools Contrasts on *Poverty* Test

	Schools					
	1 — 2		3 — 4		5 — 6	
	Effects	S.E.	Effects	S.E.	Effects	S.E.
Treatment 1	−2.00	1.36	1.03	1.34	−0.01	1.24
Treatment 2	−3.12*	1.29	3.77*	1.27	1.00	1.24
Treatment 3						

*significant at p = .05

TABLE 41

Estimated Means on Problem Task: *Goldfields* Administered at Delayed Test Session

	Schools					
	1	2	3	4	5	6
Treatment 1	4.27	2.55	1.71	2.72	1.20	2.04
Treatment 2	3.19	3.43	3.48	0.[illegible]	1.52	1.52
Treatment 3	2.08	3.56	0.78	2.48	1.44	0.61

TABLE 42

Estimates of Effects and Standard Error of Estimation Between Treatments x School Contrasts on *Goldfields* Test

	Schools					
	1 — 2		3 — 4		5 — 6	
	Effects	S.E.	Effects	S.E.	Effects	S.E.
Treatment 1	3.20*	1.43	0.69	1.42	−1.67	1.31
Treatment 2	1.25	1.37	4.41*	1.33	−0.83	1.31
Treatment 3						

*significant at $p < .05$

TABLE 43

Estimates of Effects Between Teachers on Problem Task: *Goldfields*

	Estimated Means	Estimates of Effects	Standard Error of Estimation	Estimates of Effects in S.D. units
Teacher 1	3.18	1.79*	0.39	1.12
Teacher 2	1.99	0.60	0.39	0.37
Teacher 3	1.39			

TABLE 44

Summary of Analysis of Variance on Transfer Tasks

		Simultaneous Test			Univariate F Statistic	
Source	df	Λ	F	df	*Indians*	*Famine*
Constant	1					
Treatment, eliminating constant, teacher, school, and class	2	0.81	9.54***	(4,348)	11.92***	14.16***
Teacher	2	0.95	2.49*	(4,348)	4.46*	1.86
School w teacher	3	0.96	1.05	(6,348)	0.65	1.65
Class w'school/teacher	6	0.94	0.87	(12,348)	0.27	1.55
Treatment x teacher	4	0.97	0.77	(8,348)	1.28	0.23
Treatment x school	6	0.92	1.18	(12,348)	0.65	1.70
Treatment x class	12	0.82	1.52	(24,348)	1.34	2.14*
					Mean Squares	
Within	175				4.12	3.08
Total	N=211					

***significant at p<.001
** significant at p<.01
* significant at p<.05

TABLE 45

Estimates of Effects Between Treatments on Transfer Task: *Indians*

	Estimated Means	Estimates of Effects	Standard Error of Estimation	Estimates of Effects in S.D. units
Treatment 1	3.45	1.68*	0.34	0.84
Treatment 2	2.58	0.81*	0.34	0.41
Treatment 3	1.77			

*significant at p<.05

TABLE 46

Estimates of Effects Between Treatments on Transfer Task: *Famine*

	Estimated Means	Estimates of Effects	Standard Error of Estimation	Estimates of Effects in S.D. units
Treatment 1	3.66	1.57*	0.30	0.89
Treatment 2	2.62	0.53*	0.30	0.31
Treatment 3	2.09			

*significant of $p < .05$

TABLE 47

Estimates of Effects Between Teachers on Transfer Task: *Indians*

	Estimated Means	Estimates of Effects	Standard Error of Estimation	Estimates of Effects in S.D. units
Teacher 1	3.00	0.97*	0.34	0.47
Teacher 2	2.76	0.74*	0.34	0.36
Teacher 3	2.03			

*significant at $p < .05$

TABLE 48

Summary of Analysis of Variance for Test-Retest: Control Group

		Simultaneous Test			Univariate F Statistic			
Source	df	Λ	F	df	*Poverty*	*City*	*Indians*	*Famine*
Constant	1							
Tests	1	0.98	0.30	(4,65)	0.09	0.09	1.03	0.16
					Mean Squares			
Within	68				1.88	3.81	3.32	2.96
Total	N=70							

TABLE 49

Summary of Variance for Test-Retest: Total Sample

		Simultaneous Test			Univariate F Statistic			
Source	df	Λ	F	df	*Poverty*	*City*	*Indians*	*Famine*
Constant	1							
Tests, eliminating constant, and treatment	1	0.98	0.76	(4,199)	1.28	0.70	0.20	0.52
Treatment	2	0.77	7.11***	(8,398)	10.94***	11.69***	12.34***	13.14***
Tests x treatment	2	0.97	0.74	(8,398)	1.06	0.02	1.24	1.29
					Mean Squares			
Within	202				2.74	3.79	4.24	3.40
Total	N=208							

***significant at p<.001
** significant at p<.01
* significant at p<.05

TABLE 50

Correlation Coefficients Between Component Measures of Enquiry and Problem Tasks (N=172)

	Poverty	*City*	*Indians*	*Famine*
Poverty Concept	0.14	0.22*	0.23*	0.18*
City Concept	0.18*	0.27*	0.28*	0.12
Higgins 2	0.04	0.11	0.21*	0.17*
Higgins 4	−0.03	0.10	0.15*	0.01
Answers	0.06	0.18*	0.30*	0.28*
Puzzles	0.15*	0.18*	0.40*	0.20*
Working It Out	0.08	0.21*	0.39*	0.24*

*significant at p<.05

TABLE 51

Chi-Square Tests of Successive Latent Roots for Canonical Analysis of Seven Enquiry Components and Four Problem Tasks

Number of Roots Removed	Canonical R	R Squared	Chi-Squared	df	p
0	0.53	0.28	71.1	28	<.001
1	0.23	0.05	17.5	18	N.S.
2	0.19	0.04	8.8	10	N.S.
3	0.12	0.02	2.5	4	N.S.

TABLE 52

Transformation Weights and Factor Structure Coefficients of Canonical Variates

	Transformation Weights	Structure Coefficients
Predictor Measures	I	I
Poverty Concept	0.32	0.51
City Concept	0.30	0.58
Higgins 2	0.03	0.42
Higgins 4	−0.21	0.26
Answers	0.33	0.63
Puzzles	0.35	0.74
Working It Out	0.31	0.75
Variance Extracted	0.34	
Criterion Measures		
Poverty	0.05	0.38
City	0.25	0.62
Indians	0.73	0.94
Famine	0.22	0.64
Variance Extracted	0.46	
Canonical R	0.53	
Canonical R^2	0.26	
Redundancy of Criteria given Predictors	0.13	
Total Redundancy	0.15	

References

Aebli, H. (1970). Piaget, and beyond. *Interchange, 1*, pp. 12-24.

Allen, D.I. (1970). Some effects of advance organizers and level of question on the learning and retention of written social studies material. *Journal of Educational Psychology, 61*, pp. 333-339.

Allender, J.S. (1969). A study of inquiry activity in elementary school children. *American Education Research Journal, 6*, pp. 543-558.

Anderson, R.C. (1965). Can first graders learn an advanced problem solving skill? *Journal of Educational Psychology, 56*, pp. 283-294.

Athey, J., & Rubadeau, D.O. (1970). *Educational Implication of Piaget's Theory*. Waltham, Mass: Ginn-Blaisdell.

Ausubel, D.P. (1960). The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of Educational Psychology, 51*, pp. 267-272.

Ausubel, D.P. & Robinson, F.G. (1969). *School Learning: An Introduction to Educational Psychology*. New York: Holt, Rinehart & Winston.

Banks, J.A. (1973). *Teaching Strategies for the Social Studies*. Reading, Mass: Addison Wesley.

Bartlett, F. (1958). *Thinking: An Experimental and Social Study*. New York: Basic Books.

Bayliss, D.G. & Renwick, T.M. (1966). Photograph study in a junior school. *Geography, 51*, pp. 322-329.

Beilin, H. (1971). Developmental stages and developmental processes. In D.R. Green, M.P. Ford & G.B. Flamer (eds), *Measurement and Piaget*. New York: McGraw-Hill.

Bell, D. (1973). Change and continuity: A conceptual approach to historical education. Curriculum & Research Branch, Mimeograph.

Bereiter, C. (1969). The future of individual differences. *Harvard Educational Review, 39*, pp. 310-318.

Berlak, H. (1965). The teaching of thinking. *School Review, 73*, pp. 1-13.

Blachford, K.R. (1971). What is geography in the curriculum? *Geographical Education, 1*, pp. 216-222.

Blachford, K.R. (1973). Myths in geographical education. *Geographical Education, 2*, pp. 11-25.

Blackell, B.M., Goodger, D.H., Haig, B.D., Izard, J.E., Smith, F.L., & Whitehead, G.J. (1969). *Individual Mathematics Programme*. Melbourne: Australian Council for Educational Research.

Blatt, M., & Kohlberg, L. (1974). The effects of classroom moral discussion upon children's level of moral development. Reported by J. Rest, Developmental psychology as a guide to value education: A review of 'Kohlbergian' programs. *Review of Educational Research, 44*, pp. 241-259.

Boast, K. (1973). The ability of primary and junior secondary students to read photographs. *Curriculum and Research Bulletin, 8*, pp. 2-11.

Boast, K. (1974). A study of the ability of primary school students to read photographs. *Curriculum and Research Bulletin, 9*, pp. 130-141.

Bracht, G.H., & Glass, G.V. (1968). The external validity of experiments. *American Educational Research Journal, 5*, pp. 437-467.

Brainerd, C.J. (1973). Neo-Piagetian training experiments revisited: Is there any support for the cognitive-developmental stage hypotheses? *Cognition, 2*, pp. 349-370.

Bredderman, T.A. (1973). The effect of training on the development of the ability to control variables. *Journal of Research in Science Teaching, 10*, pp. 189-200.

Bruner, J.S. (1960). *The Process of Education*. Cambridge, Mass: Harvard University Press.

Bryant, P.E. (1971). Cognitive development. *British Medical Bulletin, 27*, pp. 200-205.

Brydon, J. An investigation of certain factors entering into the judgements of adolescent boys and girls. Unpublished thesis for Diploma in Child Psychology, University of Birmingham, 1967. Reported in E.A. Peel (1971). *The Nature of Adolescent Judgment*. London: Staples.

Butts, D.P. & Jones, H.L. (1966). Inquiry training and problem solving in elementary school children. *Journal of Research in Science Teaching, 4*, pp. 21-27.

Cambourne, B.L. (1973). The verbal world of some grade I children: Implications for educators. *Victorian Institute of Educational Research Bulletin, 31*, pp. 1-13.

Case, D., & Collinson, J. (1962). The development of formal thinking in verbal comprehension. *British Journal of Educational Psychology, 32*, pp. 103-111.

Case, R. (1970). Information processing, social class, and instruction: A developmental investigation. Unpublished doctoral dissertation, University of Toronto.

Case, R. (1972*a*). Learning and development: A neo-Piagetian interpretation. *Human Development, 15*, pp. 339-358.

Case, R. (1972*b*). Validation of a neo-Piagetian capacity construct. *Journal of Experimental Child Psychology, 14*, pp. 287-302.

Case, R. (1974*a*). Structures and strictures: Some functional limitations on the course of cognitive growth. *Cognitive Psychology, 6*, pp. 544-573.

Case, R. (1974*b*). A new look at the limitations of development on the process of learning and instruction. Mimeograph.

Case, R. (1974*c*). Mental strategies, mental capacity, and instruction: A neo-Piagetian investigation. *Journal of Experimental Child Psychology, 18*, pp. 382-397.

Case, R., & Fry, C. (1973). Evaluation of an attempt to teach scientific inquiry and criticism in a working class high school. *Journal of Research in Science Teaching, 10*, pp. 135-142.

Case, R., Globerson, T. (1974). Field independence and central computing space. *Child Development, 45*, pp. 772-778.

Cleaves, R. (1972). What skills? *Occasional Papers*, Education Department of Victoria.

Connelly, F.M. (1969). Philosophy of science and the science curriculum. *Journal of Research in Science Teaching, 6*, pp. 108-113.

Cooley, W.W., & Lohnes, P.R. (1971). *Multivariate Data Analysis*. New York: Wiley.
Cordier, R.W. (1968). What should children learn through social studies? *Elementary School Journal, 69*, pp. 55-67.
Cousins, J.E. (1963). The development of reflective thinking in an eighth grade social studies class. In B.G. Massialas (ed.). *The Indiana Experiments in Inquiry: Social Studies*. Bulletin of the School of Education Indiana University, pp. 36-73.
Covington, M.V. (1968). Promoting creative thinking in the classroom. *Journal of Experimental Education, 37*, pp. 22-30.
Covington, M.V. (n.d.). Experiments to foster productive thinking in the classroom. Mimeograph.
Covington, M.V., Crutchfield, R.C., Davies, L., & Olton, R.M. (1972). *The Productive Thinking Program: A Course in Learning to Think*. Columbus: Merrill.
Cox, C.B. (1963). A description and appraisal of a reflective method of teaching United States history. In B.G. Massialas (ed.), *The Indiana Experiments in Inquiry: Social Studies*. Bulletin of the school of Education Indiana University, pp. 74-113.
Crabtree, C. (1967). Effects of structuring on the productiveness of children's thinking. *Journal of Experimental Education, 36*, pp. 1-13.
Crickmore, D., & Leighton, T. (1975). *What are the Alternatives?: Some Programs in Primary Social Studies*. Melbourne: Education Department of Victoria.
Dale, L.G. (1970). The growth of systematic thinking: Replication and analysis of Piaget's first chemical experiment. *Australian Journal of Psychology, 22*: pp. 277-286.
Darlington, R.B., Weinberg, S.H., & Walberg, H.J. (1973). Canonical variate analysis and related techniques. *Review of Educational Research, 43*, pp. 433-454.
Denney, N.W. (1974). Evidence for developmental changes in categorization criteria for children and adults. *Human Development, 17*, pp. 41-53.
Dewey, J. (1916). *Democracy and Education*. New York: Macmillan.
Dewey, J. (1933). *How We Think*. New York: Heath.
Dewey, J. (1938). *Logic: The Theory of Inquiry*. New York: Holt, Rinehart & Winston.
Donaldson, M. (1963). *A Study of Children's Thinking*. London, Tavistock.
Duckworth, E. (1972). The having of wonderful ideas. *Harvard Educational Review, 42*, pp. 217-231.
Durkin, M.C., Fraenkel, J.R., & Tanabe, P.I. (1969). *The Taba Social Studies Curriculum*. Menlo Park: Addison-Wesley.
Dwyer, F. (1971). Color as an instructional variable. *AV Communication Review, 19*, pp. 399-416.
Elsmere, R.T. (1963). An experimental study utilizing the problem-solving approach in teaching United States history. In B.G. Massialas (ed.), *The Indiana Experiments in Inquiry: Social Studies*. Bulletin of the School of Education Indiana University, p. 114.
Engelmann, S. (1969). *Conceptual Learning*. Belmont: Fearon.
Ennis, R.H. (1962). A concept of critical thinking. *Harvard Educational Review, 32*, pp. 81-111.
Evans, G.T. (1972). Some theoretical issues concerned with children's problem solving. Paper presented at the Australian Conference on Cognitive Development.

Evans, G.T., Poole, M.E., Georgeff, M., & Heffernan, G. (1974). Report to the ARGG on training in information processing strategies for children with language deficits. Mimeograph.
Fenton, E. (1966). *Teaching the Social Studies in Secondary Schools: An Inductive Approach*. New York: Holt, Rinehart & Winston.
Finn, J.D. (1972). *Multivariance: Univariate and Multivariate Analysis of Variance, Covariance, and Regression*. Ann Arbor: National Educational Resources.
Finn, J. D. (1974). *A General Model for Multivariate Analysis*. New York: Holt, Rinehart & Winston.
Finn, J.D., & Mattsson, I. (1974). *Multivariate Analysis in Educational Research: Some Applications*. Stockholm: Institute for the Study of International Problems in Education.
Fish, A.S., & Goldmark, B. (1966). Inquiry method: Three interpretations. *Science Teacher, 33*, pp. 13-15.
Flavell, J.H. (1963). *The Developmental Psychology of Jean Piaget*. Princeton: Van Nostrand.
Flavell, J.H. (1970). Concept development. In P.H. Mussen (ed.) *Carmichael's Manual of Child Psychology*, Volume 1. New York: Wiley.
Flavell, J.H. (1971). Stage-related properties of cognitive development. *Cognitive Psychology, 2*, pp. 421-453.
Flavell, J.H., & Wohwill, J.F. (1969). Formal and functional aspects of cognitive development. In D. Elkind & J.H. Flavell (eds), *Studies in Cognitive Development: Essays in Honor of Jean Piaget*. London: Oxford University Press.
Frase, L.T. (1970). Boundary conditions for mathemagenic behaviours. *Review of Educational Research, 40*, pp. 337-347.
Frase, L.T., Patrick, E.M., & Schumer, H. (1970). Effect of question position and frequency on learning from text under different levels of incentive. *Journal of Educational Psychology, 61*, pp. 52-56.
Freyberg, P.S. (1966). Stages in cognitive development–general or specific? *New Zealand Journal of Educational Studies, 1*, pp. 64-77.
Friedlander, B.Z. (1965). A psychologist's second thoughts on concepts, curiosity, and discovery in teaching and learning. *Harvard Educational Review, 35*, pp. 18-37.
Furth, H.G. (1970). *Piaget for the Teacher*. Englewood Cliffs: Prentice Hall.
Furth, H.G., & Wachs, H. (1974). *Piaget's Theory in Practice: Thinking goes to School*. New York: Oxford University Press.
Gates, A.I. & Taylor, G.A. (1925). An experimental study of the nature of improvement resulting from practice in a mental function. *Journal of Educational Psychology, 16*, pp. 583-592.
Gibson, J.J. (1954). A theory of pictorial perception. *AV Communication Review, 2*, pp. 3-23.
Goldman, R.J. (1965). The application of Piaget's schema of operational thinking to religious story data by means of the Guttman scalogram. *British Journal of Educational Psychology, 37*, pp. 158-170.
Goldschmid, M.L. (1971). The role of experience in the rate and sequence of cognitive development. In D.R. Green, M.P. Ford & G.B. Flamer (eds), *Measurement and Piaget*. New York: McGraw-Hill.
Gray, K. (1972). Thinking abilities as objectives in curriculum development. *Educational Review, 24*, pp. 237-250.
Gross, R.E. & Muessig, R.H. (eds) (1971). *Problem-centred Social Studies Instruction. Approaches to Reflective Thinking*. Washington: National Council for the Social Studies.

Haines, F. (1938). The northward spread of horses among the Plains Indians. *American Anthropologist, 40*, pp. 429-437.
Halford, G.S. (1972). The impact of Piaget on psychology in the seventies. In P.C. Dodwell (ed.), *Psychology Seventy-Two*. Harmondsworth: Penguin.
Hallam, R.N. (1967). Logical thinking in history. *Educational Review, 19*, pp. 183-202.
Harootunian, B., & Tate, M.W. (1960). The relationship of certain selected variables to problem solving ability. *Journal of Educational Psychology, 51*, pp. 326-333.
Henderson, K.B. (1972). The teaching of critical thinking. *The Educational Forum, 37*, pp. 45-52.
Henderson, W.F. (1968). The interpretation of photographs. *Curriculum and Research Bulletin, 3*, pp. 74-80.
Hermann, G. (1969). Learning by discovery: A critical review of studies. *Journal of Experimental Education, 38*, pp. 58-72.
Herron, M.D. (1971). The nature of scientific enquiry. *School Review, 79*, pp. 171-212.
Higgins, L.C. (1974). Children's inference drawing: A summarizing report of a psychological and educational study. Mimeograph.
Hills, J.L. (1970). Building and using inquiry models in the teaching of geography. In P. Bacon (ed.), *Focus on Geography: Key Concepts and Teaching Strategies*. Washington: National Council for the Social Studies.
Hirst, P. (1969). The logic of the curriculum. *Journal of Curriculum Studies, 1*, pp. 142-158.
Hirst, P., & Peters, R.S. (1970). *The Logic of Education*. London: Routledge & Kegan Paul.
Hunkins, F.P. (1970). Analysis and evaluation questions: Their effect upon critical thinking. *Educational Leadership, 3*, pp. 697-705.
Hunkins, F.P., & Shapiro P. (1967). Teaching critical thinking in elementary social studies. *Education, 88*, pp. 68-72.
Imperatore, W. (1970). On the nature of concepts. *Journal of Geography, 69*, pp. 173-178.
Inhelder, B., & Piaget, J. (1958). *The growth of Logical Thinking from childhood to Adolescence*. London: Routledge & Kegan Paul.
Inhelder, B., & Sinclair, H. (1969). Learning cognitive structures. In P.H. Mussen, J. Langer, M. Covington (eds), *Trends and Issues in Developmental Psychology*. New York: Holt, Rinehart & Winston.
Jenkinson, M.D., & Lampard, D.M. (1959). Problem solving and achievement. *Alberta Journal of Educational Research, 5*, pp. 66-172.
Johnston, C.F. (1972). An analysis of the development of logical thinking in adolescence. Unpublished B.A. (Hons) thesis, University of Sydney.
Jones, P.A. (1972). Formal operational reasoning and the use of tentative statements. *Cognitive Psychology, 3*, pp. 467-471.
Jurd, M.F. (1973). Adolescent thinking in history-type material: *Australian Journal of Education, 17*, pp. 2-17.
Kaltsounis, T. (1969). Cognitive learning and the social studies. *Educational Leadership, 26*, pp. 613-621.
Keeves, J.P. (1974). *The Home, the School and Achievement in Mathematics and Science*. Melbourne: Australian Council for Educational Research.
Klopfer, L.E. (1971). Evaluation of learning in science. In B. Bloom, T. Hastings & G. Madaus (eds), *Handbook on Formative and Summative Evaluation of Student Learning*. New York; McGraw-Hill.

Kohnstamm, G.A. (1963). An evaluation of part of Piaget's theory. *Psychologica, 21*, pp. 313-356.
Kohnstamm, G.A. (1970). Experiments on teaching Piagetian thought operations. In J. Hellmuth (ed.), *Cognitive Studies, Volume 1*. New York: Brunner-Mazel.
Koran, J.J., & Koran, M.L. (1973). Differential response to structure of advance organizers in science instruction. *Journal of Research in Science Teaching, 10*, pp. 347-353.
Kydd, S. (1969). Some insight into children's thinking background. *Curriculum and Research Bulletin, 4*, pp. 10-16.
Lodwick, A.R. (1959). An investigation of the question whether the inferences that children draw in learning history correspond to the stages of mental development that Piaget postulates. Unpublished Dip. Education dissertation, University of Birmingham, 1958. Reported in E.A. Peel, Experimental examination of some of Piaget's schemata concerning children's perception and thinking and a discussion of their educational significance. *British Journal of Educational Psychology, 29*, pp. 98-103.
Lovell, K. (1961). A follow-up study of Inhelder and Piaget's *The Growth of Logical Thinking*. *British Journal of Psychology, 52*, pp. 143-155.
Lovell, K. (1971). Some problems associated with formal thought and its assessment. In D.R. Green, M.P. Ford, & G.B. Flamer (eds), *Measurement and Piaget*. New York: McGraw-Hill, pp. 81-93.
Lunzer, E.A. (1970). Construction of a standardised battery of Piagetian tests to assess the development of effective intelligence: *Research in Education, 3*, pp. 57-72.
McLaughlin, G.H. (1963). Psycho-logic: A possible alternative to Piaget's formulation. *British Journal of Education Psychology, 33*, pp. 61-67.
McNally, D.W. (1970). The incidence of Piaget's stages of thinking as assessed by tests of verbal reasoning in several Sydney schools. *Forum of Education, 29*, pp. 124-134.
McNaughton, A.H. (1965). An analysis of concepts and content in children's responses to questions about social studies material. Paper read at ANZAAS Conference, Hobart. Mimeograph.
McNaughton, A.H. (1966). Piaget's theory and primary school social studies. *Educational Review, 19*, pp. 23-32.
Mason, J.M. (1963). The direct teaching of critical thinking in grades four through six. *Journal of Research in Science Teaching, 1*, pp. 319-328.
Massialas, B.G. (1963). Developing a method of inquiry in teaching world history. In B.G. Massialas (ed.), *The Indiana Experiments in Inquiry: Social Studies*. Bulletin of the School of Education Indiana University, pp. 1-35.
Massialas, B.G., & Cox, C.B. (1966). *Inquiry in Social Studies*. New York: McGraw-Hill.
Massialas, B.G., & Zevin, J. (1964). Teaching social studies through discovery. *Social Education, 28*, pp. 384-387.
Massialas, B.G., & Zevin, J. (1967). *Creative Encounters in the Classroom*. New York: Wiley.
Michaelis, J.U. (1973). Inquiry processes in the social sciences. Mimeograph.
Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, pp. 81-97.
Miller, J.W. (1967). Measuring perspective ability. *Journal of Geography, 66*, pp. 167-172.
Moore, D.M., & Sasse, E.B. (1971). Effect of size and type of still projected pictures on immediate recall of content. *AV Communication Review, 19*, pp. 437-450.

Moray, N. (1967). Where is capacity limited? A survey and a model. *Acta Psychologica, 27*, pp. 84-92.
Mugge, D.J. (1974). Social studies beginnings: Piagetian theory and sensitive periods. *Elementary School Journal, 74*, pp. 399-407.
Murray, F.B. (1970). Stimulus mode and the conservation of weight and number. *Journal of Educational Psychology, 61*, pp. 287-291.
Náy, M. & Associates. (1971). A process approach to teaching science. *Science Education, 55*, pp. 197-207.
Nelson, M., & Abraham, E. (1973). Inquiry skill measures. *Journal of Research in Science Teaching, 10*, pp. 291-297.
Nelson, C., & Mason, J. (1963). A test of science comprehension for upper elementary grades. *Science Education, 47*, pp. 319-327.
Nettle, E. (1975). What do children know about aborigines? Unpublished M.A. (Hons) thesis, Macquarie University.
New South Wales Education Department. (1975). *Social Studies Guidelines*. Sydney: Education Department of New South Wales.
Ojemann, R.H., & Campbell, A. (1974). The development of moral judgments I. *Journal of Experimental Education, 42*, pp. 65-66.
Olton, P.K., Raskin, E., & Witkin, H.A. (1971). *Group Embedded Figures Test*. Palo Alto: Consulting Psychologists Press.
Olton, R.M., & Crutchfield, R.C. (1969). Developing the skills of productive thinking. In P.H. Mussen, J. Langer, M. Covington (eds), *Trends and Issues in Developmental Psychology*. New York: Holt, Rinehart & Winston.
Orlandi, L.R. (1971). Evaluation of learning in secondary school social studies. In B. Bloom, T. Hastings & G. Madaus (eds), *Handbook on Formative and Summative Evaluation of Student Learning*. New York: McGraw-Hill.
Palfrey, C.F. (1972). Piaget's questions to young children: An analysis of their structure and content. *Educational Review, 24*, pp. 122-131.
Parkinson, G.M. (1969). The recognition of messages from visual compound stimuli: A test of a quantitative developmental model. Unpublished Masters thesis, New York University.
Parnes, S.J., & Meadow, A. (1959). Effects of brainstorming instruction on creative problem-solving by trained and untrained subjects. *Journal of Educational Psychology, 50*, pp. 171-176.
Pascual-Leone, J. (1970). A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologica, 63*, pp. 301-345.
Pascual-Leone, J., & Smith, J. (1969). The encoding and decoding of symbols by children: A new experimental paradigm and a neo-Piagetian model. *Journal of Experimental Child Psychology, 8*, pp. 328-355.
Pate, R.T., & Bremer, N.H. (1967). Guiding learning through skilful questioning. *Elementary School Journal, 67*, pp. 417-422.
Peel, E.A. (1966). A study of differences in the judgments of adolescent pupils. *British Journal of Educational Psychology, 36*, pp. 77-86.
Peel, E.A. (1971). *The Nature of Adolescent Judgment*. London: Staples.
Piaget, J. (1964). Cognitive development in children: The Piaget papers. In R.E. Ripple & V. Rockcastle (eds), *Piaget Rediscovered: A Report of the Conference on Cognitive Studies and Curriculum Development*. Ithaca, New York: Cornell University.
Piaget, J. (1968). Quantification, conservation and nativism. *Science, 162*, pp. 967-969.
Piaget, J. (1970). Piaget's theory. In P.H. Mussen (ed.), *Carmichael's Manual of Child Psychology, Volume 1*. New York: Wiley.

Piaget, J. (1971a). Commenting on Beilin's paper. In D.R. Green, M. Ford & G.B. Flamer (eds), *Measurement and Piaget*. New York: McGraw-Hill.
Piaget, J. (1971b). *Science of Education and the Psychology of the Child*. London: Longman.
Piaget, J. (1972). Intellectual evolution from adolescence to adulthood. *Human Development, 15*, pp. 1-12.
Piper, M.K. (1969). A study of the effect of teaching productive thinking skills to fifth-grade students using The Productive Thinking Program. M.A. thesis, University of Kansas.
Pledger, P. (1972). Student maturity and the consideration of controversial issues. Mimeograph.
Poole, M.E. (1973). Linguistic, cognitive and verbal processing styles: A social class contrast. Unpublished doctoral dissertation, La Trobe University.
Posner, M.I. (1965). Memory and thought in human intellectual performance. *British Journal of Psychology, 56*, pp. 196-215.
Pryor, L.J. (1943). History in the elementary school. *Educational Gazette and Teachers' Aid, 43*, pp. 80-82.
Queensland, Education Department. (1972). *A Guide for the Social Studies Teacher*. Brisbane: Education Department of Queensland.
Renehan, W.T., & Wilkes, R.E. (1973). *Report on Victorian Primary Schools Testing Programs 1969-1970*. Hawthorn: Australian Council for Educational Research.
Rest, J. (1974). Developmental psychology as a guide to value education: A review of 'Kohlbergian' programs. *Review of Educational Research, 44*, pp. 241-259.
Rhys, W.T. (1966). The development of logical thought in the adolescent with reference to the teaching of geography in the secondary school. Unpublished M.Ed. thesis, University of Birmingham.
Ripple, R.E., & Dacey, J. (1967). The facilitation of problem solving and verbal creativity by exposure to programmed instruction. *Psychology in the Schools, 4*, pp. 240-245.
Robinson, F.G. (1973). An abridged program of inquiry skills training built around a decision-making model. Mimeograph.
Robinson, F.G. (n.d.) Tentative marking scheme. Mimeograph.
Robinson, F.G., Tickle, J., & Brison, D.W. (1972). *Inquiry Training: Fusing Theory and Practice*. Profiles in Practical Education No. 4. Ontario Institute for Studies in Education.
Rothkopf, E.Z. (1966). Learning from written instructive material: An exploration of the control of inspection behavior by testlike events. *American Educational Research Journal, 3*, pp. 241-249.
Rothkopf, E.Z. (1970). The concept of mathemagenic activities. *Review of Educational Research, 40*, pp. 325-336.
Ryan, F.L. (1973). Differentiated effects of levels of questions on student achievement. *Journal of Experimental Education, 41*, pp. 63-67.
Saarni, C.I. (1973). Piagetian operations and field independence as a factor in children's problem solving performance. *Child Development, 44*, pp. 338-345.
Schwab, J.J. (1964). Structure of the disciplines: Meanings and significances. In G.W. Ford & L. Pugno (eds), *The Structure of Knowledge and the Curriculum*. Chicago: Rand McNally.
Schwebel, M., & Raph, J. (1974). *Piaget in the Classroom*. London: Routledge & Kegan Paul.
Scott, N. (1973). Cognitive style and inquiry strategy: A five year study. *Journal of Research in Science Teaching, 10*, pp. 323-330.

Shaver, J.P., & Larkins, A.G. (1973). Research on teaching social studies. In R.M. Travers (ed.), *Second Handbook of Research on Teaching*. Chicago: Rand McNally.
Shulman, L.S. (1965). Seeking styles and individual differences in patterns of inquiry. *School Review, 73*, pp. 258-266.
Shulman, L.S., & Keislar, K.R. (eds) (1966). *Learning by Discovery: A Critical Appraisal*. Chicago: Rand McNally.
Siegel, L.S. (1968). The development of the ability to process information. *Journal of Experimental Child Psychology, 6*, pp. 368-383.
Sigel, I.E. (1969). The Piagetian system and the world of education. In D. Elkind & J. H. Flavell (eds), *Studies in Cognitive Development: Essays in Honor of Jean Piaget*. London: Oxford University Press.
Sigel, I.E. (1974). When do we know what a child knows. *Human Development, 17*, pp. 201-217.
Simons, M. (1971). Geographical concepts and education. *Geographical Education, 1*, pp. 207-215.
Sinclair, H. (1971). Representation and memory. In M.F. Rosskopf, L.D. Steffe & S. Taback (eds.), *Piagetian Cognitive: Development Research and Mathematical Education* Washington: National Council of Teachers of Mathematics.
Sinclair, K.E. (1973). Some correlates of cognitive development during adolescence. Mimeograph.
Smedslund, J. (1966). Performance on measurement and pseudomeasurement tasks by five- to seven-year-old children. *Scandinavian Journal of Psychology, 7*, pp. 81-92
Smith, B.D. (1957). Logic, thinking and teaching. *Educational Theory, 7*, pp. 225-233
Snow, R.E. (1974). Representative and quasi-representative designs for research on teaching. *Review of Educational Research, 44*, pp. 265-291.
Suchman, J.R. (1961). Inquiry training: Building skills for autonomous discovery. *Merrill-Palmer Quarterly of Behavior & Development, 7*, pp. 147-169.
Suchman, J.R. (1964). The Illinois studies in inquiry training. *Journal of Research in Science Teaching, 2*, pp. 230-232.
Suchman, J.R. (1966). A model for the analysis of inquiry. In H.J. Klausmeier & C.W. Harris (eds), *Analyses of Concept Learning*. New York: Academic Press.
Sullivan, E.V. (1969). Piagetian theory in the educational milieu: A critical appraisal. *Canadian Journal Behavioral Sciences, 1*, pp. 129-153.
Taba, H., Levine, S., & Elzey, F. (1964). *Thinking in Elementary School Children*. San Francisco: U.S. Office of Education.
Taba, H. (1966) *Teaching Strategies and Cognitive Functioning in Elementary School Children*. San Francisco: U.S. Department of Health, Education & Welfare.
Taba, H., Durkin, M., Fraenkel, J., & McNaughton, A. (1971). *A Teacher's Handbook to Elementary Social Studies: An Inductive Approach*. Reading: Addison-Wesley.
Tanner, J.M. (1960). Comments on Professor Piaget's paper. In J.M. Tanner & B. Inhelder (eds), *Discussions on Child Development: Volume IV*. London: Tavistock.
Tomlinson, P.D. (1971). An investigation into the development of short-term retention capacity in children as a function of age and item familiarity. Unpublished doctoral dissertation. University of Toronto.
Traub, R.E., Weiss, J., Fisher, G.W., & Musella, D. (1972). Closure on openness: Describing and quantifying open education. *Interchange, 3*, pp. 69-84.

Travers, M.W. (1969). A study of the advantages and disadvantages of using simplified visual presentations in instructional materials. Final Report on USA-Grant No. 1-7-070144-5235, cited in Travers, M.W. & Alvarado, V. (1970).

Treffinger, D., & Ripple, R. (1970). The effect of programmed instruction on creative problem solving and attitudes. *Irish Journal of Education, 1*, pp. 47-59.

Trezise, R.L. (1974). Developing performance objectives in social studies in Michigan. *Social Education, 38*, pp. 24-29.

Tunstall, K.W. (1969). Computer assisted group problem solving. Unpublished doctoral dissertation, University of Toronto.

Vernon, M.D. (1968). Perception and perceptual learning. In E.A. Lupzer & J.F. Morris (eds). *Development in Human Learning*. London: Staples.

Victoria, Education Department. (1968*a*). *Teachers' Guide, History: The Use of Sources*. Melbourne: Education Department of Victoria.

Victoria, Education Department. (1968*b*). *Curriculum Guide Mathematics*. Melbourne: Education Department of Victoria.

Victoria, Education Department. (1969). *Primary Science Curriculum Guide: Branching Out*. Melbourne: Education Department of Victoria.

Victoria, Education Department. (1974). *Society in View: A Process Approach to Social Competence*. Melbourne: Education Department of Victoria.

Wallen, N.E., Durkin, M.C., Fraenkel, J.R., McNaughton, A.J., & Sawin, E.I. (1969). *The Taba Curriculum Project in Social Studies*. San Francisco: U.S. Department of [redacted] Education & Welfare.

Wardrop, J.L., Goodwin, W.L., Klausmeier, H.J., Olton, R.M., Covington, M.V., & Crutchfield, R.S. (1969). The development of productive thinking skills in fifth-grade children. *Journal of Experimental Education, 37*, pp. 67-77.

Western Australia, Education Department. (1974). *Social Studies*. Perth: Education Department of Western Australia.

Whitehead, G.J. (1971). New lamps for old or old lamps reburnished: The new social studies. *Victorian Institute of Educational Research Bulletin, 26*, pp. 12-25.

Whitehead, G.J. (1972). The cognitive levels of grade VI, form II and form IV students when solving social studies type problems and the influence upon performance of varying the method of question presentation. Unpublished M.Ed. thesis, University of Melbourne.

Whitehead, G.J. (1973). The Eskimos are back: Man, a course of study. *Curriculum and Research Bulletin, 8*, pp. 85-92.

Whitehead, G.J. (1975). Cognitive performance and social studies: The impact of two training programs on grade six students. Unpublished doctoral dissertation, La Trobe University.

Wicks, J.M. (1974). The generation of alternatives and thinking styles in elementary social studies. Unpublished doctoral dissertation, University of Alberta.

Witkin, H.A., Dyk, R.B., Faterson, H.F., Goodenough, D.R., & Karp, S.A. (1962). *Psychological Differentiation*. New York: Wiley.

Witkin, H.A., Goodenough, D.R., & Karp, S.A. (1967). Stability of cognitive style from childhood to young adulthood. *Journal of Personality and Social Psychology, 7*, pp. 291-300.

Programs in primary school social studies advocate the development and application of enquiry skills but children seem to have difficulty carrying out their investigations. Why should this be? What are the actual skills involved and why should children have difficulty in applying them?

Enquiry Learning in Social Studies examines these issues, and describes and assesses the adequacy of a program which attempts to match some of the objectives of enquiry in social studies with the abilities of the students being taught. Using multivariate analysis of variance Graham Whitehead reports upon the success achieved by Grade 6 students in solving social studies-type problems and examines whether this success can be retained and applied to other tasks after completion of the initial training program.

AUSTRALIAN COUNCIL FOR EDUCATIONAL RESEARCH ISBN 0 85563 164 3